UNIVERSITY of LIMERICK

OLLSCOIL LUIMNIGH

An Investigation of News Analytics Data and the Impact it has Upon Volatility

Jonathan Nolan

16071514

MSc in Computational Finance

2017 - 2018

Supervisor: John Garvey

Word Count: 10,406

This dissertation is solely the work of the author and submitted in partial fulfilment of the requirements of the MSc in Computational Finance.

Declaration

I hereby declare that all the content contained in this dissertation is my own unless otherwise stated. All work has been carried out in accordance of the Kemmy Business School Masters Programmes guidance. Where other work has been included, these have been correctly sourced and referenced. Any views expressed in this dissertation are those of the author alone

Signed:____________________ Date:____________________

Abstract

The research depicted in this study focuses on examining stock volatility and its dynamics, through various models. It also describes news analytics; how stock reacts to news reports and could potentially be incorporated in to forecast models to intensify the model's predictability power. The sentiment of each news source is assigned as a numerical value in order to incorporate it into mathematical models such as Pearson Correlations for analysis. The wordlists which have been used to generate the numerical values of sentiment, are examined against each other, to gain a greater understanding of their analysis results. Additionally, the frequency of information releases is examined and implemented in Aug-GARCH models. Overall, the main focus will be to study the influence of news reports upon stock values, in a financial context.

Acknowledgements

I would like to express my gratitude to my supervisor, Dr John Garvey for his guidance, direction and advice throughout the process of this research.

I would also like to show my appreciation to Dr Bernard Murphy for dedicating so much effort to the Computational Finance course which has allowed me to develop the ability to undertake such a research.

I also wish to thank my classmates for their support and guidance throughout the duration of the course.

Finally, I am very grateful to my family who have supported me always, not least throughout this past year. To my grandparents Thomas and Gretta, my parents Michael, Mary, and my brother and sisters, Amy, Lily, Emily and Mike - I am most appreciative for your time. Without their support and help, this research would not have been possible.

Table of Contents

Figures

Tables

List of Abbreviations

EMH	Efficient market hypothesis
UIH	Uncertain Information Hypothesis
LM	Loughran and Mcdonald (wordlist)
D7	Diction 7 (wordlist)
H6	Henry 2006 (wordlist)
H8	Henry 2008 (wordlist)
LIWC	Linguistic Inquiry and Word Count (wordlist)
FTSE	Financial Times Stock Exchange
ARCH	Autoregressive conditional heteroskedasticity
GARCH	Generalised autoregressive conditional heteroskedasticity
AUG-GARCH	Augmented GARCH
QMLE	Quasi Maximum Likelihood Estimates
IID	Independent and identically distributed
MDH	Mixture of Distribution Hypothesis
SEC	U.S. Securities and Exchange Commission

Chapter 1: Overview

1.1 Introduction to News

1.1.1 News within the Financial Sector

News is shared information which is widely available and accessible to the public. In the financial sector, news can span a wide variety of areas; reporting on industries, companies, regulations and more. News generally reports on significant events and is circulated as shared information of importance which enhances one's understanding about the international stock markets. The distribution of news can result in recipients of the information, interpreting the news and enabling investors to make assessments and take actions which may contribute to their portfolio of assets. It is of interest to this research project to examine the subsequent reaction of the market participants, to the distribution of news and the sentiment formed from the consequential interpretation.

The actions foreshadowing the public circulation of news broadcasts, are essential in bolstering the news stories legitimacy and reliability. This process generally involves collecting information directly from credible sources, which allows journalists to produce a drafted news piece which is then processed by an editor before being approved for circulation. This detailed reviewing process assures that only authenticated sources of news are included, thus preserving the reader's trust and confidence in the news. Additionally, this also potentially indicates that there is a filtering process which occurs that determines whether a source of information is satisfactory enough to broadcast.

The rationale behind this is as follows: the knowledge and integrity of the information is restricted to a small group of individuals initially, whom subsequently release it to journalists and media sources. It is then determined whether it is a necessity for the information to be publicised. Thus, in order to distinguish between what information is acceptable for circulation, and what information should be refrained from distribution, selection criteria are imposed. Namely, news is information which has been graded as significant for the audience of market participants such as investors, traders etc. to receive. This raises questions about the suitability of those involved in deciding the relevance of news information. At present, those engaged in the information selection process, are journalists. However, after further reflection, it is the source that distributed the information to the journalist, who decides what information to deliver in the first instance. Therefore, there must be an underlying motive which drives the source to dispense certain information and ensure that it is distributed to the public. If there was a distribution of false or unsound information, it would most likely have been a result of the person releasing the information, in an attempt to manipulate the market and generate financial gain (Kogan *et al.* 2017). However, this is an illegal act. The objective of journalists is to entice the public's attention and capture their interest. This may result in a prevailing opinion being formed amongst market participants; otherwise known as 'sentiment'.

1.1.2 The Influence of News

The revolutionary research of Bachelier *et al.* (2006), which suggests that it is not possible to forecast future stock prices by analysing historical stock prices, has become common knowledge. However, it is universally accepted that a major contributor to the process in which the price of financial assets adapts, is information flow. This is particularly relevant to a stock's trading volume and volatility, as information flow plays an influential part in the theoretical model of realised volatility (Brooks, 1998). Nonetheless, it is uncertain whether or not, new information assists in explaining or forecasting the value of stocks. The rationale behind this uncertainty, is the Efficient Market Hypothesis (EMH). The work of Fama (1970), formulated the EMH, which proposes that any new information which is being circulated, is reflected in the financial market, in an efficient manner, by providing fair market prices.

From the EMH, it is theorized that no participant in the market, can earn excess returns as all information should be precisely reflected in the price of the shares. As modern electronic markets are becoming continuously faster at interpreting and analysing new information, this should hold true. Theoretically, as all stock prices are claimed to incorporate all prior public information, investors and traders should not be able to initiate a trading strategy that will allow them to gain excess returns from analysing historical information. Therefore, according to the EMH, there should be no correlation between present-day stock returns and the sentiment of financial articles concentrating on a particular stock.

Interestingly, the work of Malkiel (2003) provides empirical evidence that in reality the EMH, does not hold true. The need for an alternative to the EMH, lead to Brown *et al.* (1988), developing the Uncertain Information Hypothesis (UIH). The overreaction hypothesis by De Bondt and Thaler (1985) provides the foundation for the UIH, which in comparison to the EMH, differs as new prices are set by market participants ahead of the full extent of the news content being determined. In a situation when there is both favourable and unfavourable news, the stock prices are set at a price which is considerably below the conditional expected values, by the investors and therefore, reacts risk-aversely. Evidence of this overreaction hypothesis is provided by De Bondt and Thaler (1985). Their research indicates that portfolios which consist of "loser" stocks, are found to outperform portfolios made up of "winners", by approximately 25%.

Sentiment within the news dataset studied by De Bondt and Thaler (1985), is not analysed or relied upon. Instead, their work defines a significant movement in the price of a stock as a "news event". The work of Zarowin (1990) provides the argument that due to size effect of the reduced market capitalization of the stocks which are labelled losers, the overreaction phenomenon is caused. It is reported in the work of Banz (1981), that on a consistent historical basis, firms who have a lesser market capitalization exceed the performance of stocks which have a larger market capitalization. Further evidence of the UIH is researched in various other works such as Pantzalis *et al.* (2000) and Yu *et al.* (2009).

All of the aforementioned research, rely on deriving news measures directly from the reactions of stocks, specifically returns. Work which has specifically concentrated on news sentiment, does not re-examine or consider the UIH theory. In the present day 'Big Data' era, not only has the volume of news articles published increased tremendously, but there is also an expanding body of literature that recognises the importance of news releases about companies, and the impact they have upon the companies' stocks (Braun *et al.* 1995; Edmans *et al.* 2014; Alfano *et al.* 2015). Due to this large amount of information which is publicly available, and the existence of computers which are capable of operating with large volumes of data, it is only reasonable to attempt to use news and other information broadcasting sources to enhance our knowledge of how stock reacts. The stock reaction which will be examined specifically in this research will be volatility, through the use of trading volume, number of news releases and returns.

One possible benefit of analysing the above-specified data using high powered advanced computers, is that they will provide insightful forecasts of a company's stock which would allow investors to filter noises and implement more calculated decisions. Two major known sources of information which investors use to extract and analyse market information, are financial news articles and press releases. The dataset upon which this thesis concentrates on, contains both press releases and media articles reporting on a range of companies, across various industries. Both the press releases and the media articles concentrate on a particular company, describe the company's performance and provide news information regarding it. The sentiment of these press releases and media articles, has been analysed and converted into numerical values which make it possible to analyse it using mathematical models of analysis.

To summarise, ordinarily a company will release information about its performance to the general public. From these releases, the media will depict the information and then construct an article based on their interpretation, reporting their analysis of the information. It is important to note however, that any large variation between the sentiments of both sources, could potentially cause uncertainty to rise within the market. This may result in the realised volatility reacting drastically due to a discrepancy between reality and what is reported by the media. Despite the fact that press releases about earning are voluntary, companies are more inclined to release information to the media and the public if they contain positive news, in an attempt to portray a prosperous quarter (Peck and Hall, 2013).

The writing style implemented, and the language used in the delivery of each earning press release, could influence media reactions. Depending on how content is delivered, a successful quarter could be maximised, or a company’s underperformance could be minimised. This may influence media reports (Maat, 2007). Furthermore, this research is interested in investigating if there are considerable discrepancies between the press releases of companies and the media articles which follow them, in terms of sentiment.

1.2 Thesis goal

As there is an ever-expanding amount of information available to the public through the internet, media and news, it is of interest to this study to utilize this wealth of information sources to improve our understanding of how stocks react to information releases. To elaborate, this reaction is the volatility of the stock.

This thesis will also expand further the analysis of sources to distinguish if the type of source releasing the information carries weight on the sentiment of the release. The sentiment which has been gathered from the various sources has been converted to numerical values through various wordlists, thus making it possible to incorporate it into mathematical models such as Pearson Correlations for analysis. Through analysis of the derived sentiment and also the volume of releases published, answers to the following research questions will be attempted:

(i) Does a positive correlation exist between the numerical tone values of Press Releases and Media Articles?
(ii) Does the type of source releasing the information have an effect on the sentiment of the release?
(iii) Does the volume of information being released, cause reactions within the market?
(iv) Can forecasts be made based on the potential evidence gathered?

Question (i) attempts to investigate if the information published by the corporate disclosures are depicted with the same numerical values of sentiment, as the media articles which are interpreting this information. Question (ii) aims to investigate if the type of source, for example, media articles and press releases, plays a role in the sentiment of the information being positive or negative. It will study if press releases are more positive in their nature when releasing information. Additionally, Question (iii) aims at an investigation into if the volume of information released has an impact upon the stock markets. Finally, Question (iv) aims to investigate if there is evidence of accurate forecasts, and if traders can benefit from this information.

To answer these questions, the investigation is split into two parts. Part 1 concentrates on the sentiment dataset which investigates Questions (i) and (ii), and is contained in chapters 3 and 4. Whereas Part 2 will attempt to answer Questions (iii) and (iv). This part of the study consists of chapters 5 and 6 and focuses on implementing the volume of news releases into forecast models.

1.3 Thesis outline

In the opening chapter, the focus of the research has been presented and outlined. Additionally, the chapter also introduced the perception of news as an ‘event’. Further research within this thesis is structured as follows. Chapter 2 will provide a discussion on previous literature of news sentiment and how it has grown over time, particularly looking at the importance of text analysis and text classification. Chapter 3 consists of the sentiment and tone theory which will be used in the investigation undertaken by this study. It will include a detailed summary of what news analytics is and also the methodology undertaken for this study. Chapter 4 provides a summary of the data set which the thesis focuses on as well as considerations to be accounted for. Furthermore, it implements the theory outlined in Chapter 3 and reports the results found. Outlined in Chapter 5 is the investigation of using the number of new publications within a GARCH(1,1) model. It includes the methodology which will be undertaken for this part of the study. Chapter 6 summarises the dataset which will be used in this investigation as well as the results obtained from the methodology discussed in Chapter 5. Finally, Chapter 7 discusses the results obtained in chapter 4 and 6. The chapter concludes with the main points to be considered in this paper.

Chapter 2: Literature review

Throughout the research and analysis of sentiment and news, the traditional sources of information have been newspaper reports such as articles from the *Wall Street Journal*. Many previous studies have investigated the relationship between the flow of news and stock reactions, however, many of them have neglected to incorporate a sentiment classification into their research. The work of Tetlock (2007), provides evidence that negative sentiment derived from the articles of a *Wall Street Journal* column, had an explanatory influence on the *Dow Jones* downward trajectory. Groß-Klußmann and Hautsch (2011), focused on analysing how the market reacts to the intra-day stock specific data. The data was provided from the *Reuters* "News Scope Sentiment" engine which had derived the numerical sentiment values of the data. Their results endorse the hypothesis that volatility and trading volume is influenced by news. However, due to their high-frequency context, it is limited to a select number of assets. Wisniewski and Lambe's (2013) study of the role of media during the credit crunch, used the *Lexis-Nexis* database as their source to collect news. From their collected information, they filtered for phrases and expressions that were notably used during the 2008 global financial crisis, for example, "Financial Crisis". The evidence gathered suggests that news may influence the future movement of the market, however, there is not substantial evidence that journalists repeat prior news.

Early research which focuses on the application of social media sentiment values on stock markets, tends to use message boards as their source of text, for example, *Yahoo! Finance*. The benefits of using these boards as a data source, are the frequency of new messages and the identifiability of the stocks which are focused on within the text. Das and Chen (2007), use the previously mentioned message board, *Yahoo! Finance*, to provide evidence that there is a positive correlation between sentiment which has been accumulated from the message board, and the return of the stocks index, the following day. However, there is no evidence provided to prove this correlation on an individual firm level. The work of Antweiler and Frank (2004) also uses *Yahoo! Finance* as their text source as well as another message board, *Ragin Bull.* Their study gathers evidence that the volume of messages, as well as their 'bullishness', has predictive value in regards to volatility. Zhang and Swanson (2009) analysed from message boards, the self-disclosed sentiment of messages which focus on holding a stock position. They concluded that they are not bias-free as they are significantly more optimistic than neutral.

Further work which has also used message boards is the study by Sabherwal *et al.* (2011). This research examined if the stock market was manipulated by focusing message boards on small-cap firms, which text contained *"pump and dump"* strategies. They were able to discover a pattern which implies there is a possibility that an online discussion has the potential to manipulate the stock prices of a small firm. Moreover, Park *et al.* (2013) investigate the effect that stock message boards have upon investors. Their results suggest that a confirmation bias exists between them, thus implying that investors favour information that echoes their prior judgements. One notable disadvantage for including message board data, is that the size of the amassed dataset samples is rather restricted. A possible reason for this could be that instead of displaying the whole history, *Yahoo! Finance* only displays a fixed quantity of messages and thus, disregards other messages provided. This is evident in the work of Antweiler and Frank (2004) and Sabherwal *et al.* (2011), who are only able to incorporate a maximum time period of one year of message posting.

Bettman *et al.* (2010) analysed a wider sample of text, which spans a time period of six years. Their sample data was filtered through by Naïve Bayes classification, for rumours involving potential takeovers. Subsequently, this research realised that at times, returns and trading volumes which were abnormal, followed the posting of these rumours. Kim and Kim (2014) is another piece of research which focuses on evaluating six years of message postings. Unlike Sabherwal *et al.* (2011), their study concludes that the prior performances of stock prices influences future message board postings rather than postings impacting the stock price performance. Further research which concentrates on message board postings is by Li *et al.* (2014), who derive news from Chinese web pages, and provide an indication that the CSI index is outperformed by their *"electronic-media-aware quantitative trader"*.

More recently, researchers have begun to test the sentiment of the social media platform *Twitter* into their studies. The message postings (tweets) that are published on this platform have a restriction of only 140 characters per post, as they tend to be written via a mobile cellular device. As a result of this restriction, the usual hindrance of multiple stocks being mentioned in a data source, as well as grammar mistakes, tend not to be evident in the analysis. Bollen *et al.* (2011) use a classification of various different mood states to separate the tweets, and their results indicate that the mood of the public can enhance the prediction power to forecast the daily changes in the *Dow Jones* index. This classification process is refined by Zhang *et al.* (2012), who focus on isolating keywords that tend to indicate a financial context. The work of Zhang *et al.* (2012) also expands their market study by including currencies and commodities.

The process of classifying was extended by Si *et al*. (2013), who filtered *Twitter* posts to collect tweets on a firm-specific level. Their study deduces that the accuracy of day-to-day predictions of stock is enhanced from the extracted sentiment that is topic specific. Another study who used tweets on a stock level is Sprenger *et al.* (2014), whose work implied that the number of followers of the tweets poster, and the volume of retweets the post receives, can successfully assess the quality of the advice within the tweet. The sentiment trading model developed by Nann *et al.* (2013) outperforms the *S&P 500*, by aggregating data from sources such as traditional newspapers, stock message boards and *Twitter*.

In more recent studies, investment communities on social media, such as *Seeking Alpha*, have been analysed. This includes not only the articles published on the community, but also the comments within the article posting. An example of a study who used this type of source is Chen *et al.* (2014), who collected data from *Seeking Alpha* and used their analysis to indicate that negative sentiment for stock returns, holds forecasting value by enhancing prediction models. *Seeking Alpha* was also used as a source by the work of Wang *et al.* (2014), which claims to have found that the correlation between the sentiment of *Seeking alpha* and returns is greater than the correlation between the sentiment of another finance blogging platform, *StockTwits,* and returns. *NASDAQ Community* is a platform which collects and accumulates investment articles from various other social media platforms and news articles. This platform was used by Zhang *et al.* (2015) in their research. Their study made a comparison of the different sentiment lexica, and found an incremental influence of the derived sentiment on the volatility and returns of a stock.

Ahern and Sosyura (2014) studied the manner in which companies portray their releases of earnings by examining the tone of the releases. The main component of their study is the abnormal positive tone and they concluded that there is evidence of firms implementing strategic disclosures in their releases. This discrepancy in tone may be due to media articles which are published on an earnings press release, potentially providing an over or under-representation of the information disclosed, or also providing a more accurate interpretation on the news released. Nonetheless, this discrepancy between the tones of the sources may result in confusion in the market, hence why they should be examined separately. This is important with regard to noise traders who could be influenced in their investment decisions by the sentiment of these sources.

Chapter 3: News Analytics

3.1 Introduction

There are numerous investment companies in the world that have been incorporating news analytics into their studies, to enhance their businesses. The underlying rationale behind this interest, is to use News Analytics to strengthen their prediction abilities of the stock market (Tetlock, 2007). Various methodologies and data mining approaches are implemented into News Analytics (Kantardzic, 2011), which include techniques from Machine Learning, Mathematical Statistics and Modelling, and Artificial Intelligence. Traders are signalled about the events in the news which may have the greatest impact on the stock market through News Analytics software. Another benefit of this software is that it can take the signals into account automatically and incorporate them into the relevant trading algorithms.

This chapter will provide a short overview of the tools, techniques used in News Analytics and the different providers of News Analytics.

3.2 What is News Analytics?

The research of Das (2012) defines News Analytics as *“the measurement of the various qualitative and quantitative attributes of textual news stories”*. Examples of these attributes are as follows:

The Sentiment of News
This distinguishes if the news is positive or negative, which could possibly have a mirrored impact on the stocks. A possible scenario is that positive news will result in an increase in a company’s stock prices, whilst negative will result in a drop (Li et al. 2011).

The Impact of News
This is characterised by the significance the News has on the stock, in the form of the scale of changes, which occur due to the News.

The Relevance
This illustrates how the events that are reported in the news, are applicable to the particular security that a trader has invested in.

The Novelty
This describes how informative the news is. Generally, to calculate the novelty, the news is inversely correlated with the number of references to events, that in turn, are included in the particular report with other news publications.

The evaluation of news as an event is a relatively fresh and new approach, which is used to enhance the trading strategies of investment funds. The behavioural finance theory is closely associated with News Analytics, whilst News Analytics in some ways, conflicts with the typical economic theory.

However, as mentioned earlier, the Efficient market hypothesis (EMH) suggests that any newly available information is already reflected in the financial market (Fama, 1970). Due to this concept, it is futile to attempt to surpass the markets’ performance over long stretches, using information which is easily available on the market. Nonetheless, it is still worth noting that a trader on their own, is incapable of managing the information flow from all news agencies, as the frequency of news reports is so great. Due to this great flow, incidents which could potentially influence the stock exchange, may not be taken in to account. Therefore, it would be rare that all investors would be uniformly informed of all financial happenings at a particular moment in time, that could impact the value of various stocks. Hence lies the importance of News Analytics as a practical tool to gain an advantage over the competition in the market.

By understanding and obtaining the components of news in numerical format, they can be implemented into statistical models and trading systems.

According to Mitra and Mitra, (2011), the automated process involved in the analysis of news in information systems, usually consists of the following procedures:

1. Compiling news reports from various sources
2. Undergoing an exploratory investigation of the collected news
3. Examination of the news sentiments, incorporating present market positions
4. Construction and implementation of quantitative models

Further detailed descriptions of News Analysis and its process are found in sections which follow.

It may be worth noting that due to managers of investment funds creating their portfolios to increase in value over long periods of time, they rarely implement tools of News Analytics.

3.3 Data Sources

There are numerous different sources where News data can be obtained, such as:

New Agencies
News used to only be limited to printed sources, radio and television, which resulted in it being challenging to attain an accurate comprehensive picture of the news flow. However, due to the aid of the internet, the process of news analysis has changed, with the inclusion of tagging and indexing of news sources allowing the process to become automatic.

Pre-News
This is a source of raw information material which is analysed and studied by reports when preparing news sources. It can be collected from various primary outlets, such as company quarterly releases, announcement and SEC reports.

Social Media Sources
Examples of social networks include *Twitter, Facebook* and various blogs. The authenticity of the news from these sources can range highly, and information collected can often be worthless. However, if a system can analyse the sentiment of a great volume of these messages, the results can be applicable in trading strategies.

Additionally, the financially related news can be evaluated by considering expectations. Due to quarterly company announcements occurring at allocated times, the expected news which reports on these announcements, can be forecasted on the evidence of pre-news. This is due to the pre-news having a structured format and in general containing numerical information which can be incorporated for an automated analysis. Reports which contain the losses and profits of a corporation, directly affect the stock prices to change in price and value, hence why they are universally included in investment strategies. The fundamental complications of processing economic information are related to the occurrence of unexpected news, as generally, they can occur at any given moment, and commonly do not include numerical data or a structured text format. Although they contain key information on the event, the aforementioned reasons prevent them from being analysed efficiently and promptly.

3.4 Providers of News Analytics

Presently, more than 50 providers exist that act as sources of economic news, the three most dominant being *Bloomberg, Dow Jones* and *Pearson*. The provider of News Analytics and data implemented in this study are:

Factiva (www.Factiva.com)

This is a current international business news database, that has access to over 36,000 different sources all over the world. This includes over 600 various new wires, which are updated frequently. An advantage of *Factiva* is it allows the user the ability to determine which articles are firm-initiated and which articles are press-initiated, due to their source. The table below displays an extract sample of the data from *Factiva*.

Variable	Description
ID	276
ISIN	GB0009895292
Company Name	ASTRAZENECA
Keyword	AstraZeneca
Type	319608
subjectsSN	5
$LP + TD(10\%)$	2
$LP + TD(20\%)$	3
$LP + TD(100\%)$	5
Section (SE)	#NA
Headline (HD)	Losec lifts drugs firm profits but shares slip - AstraZeneca.
Author (BY)	Carl Mortished International Business Editor.
Word Count (WC)	262
Publication Date (PD)	04-May-00
Publication Time (ET)	#NA
Source Name (SN)	The Times
Source Code (SC)	T
Language (LA)	English
Contact (CT)	#NA
Subject Code: Descriptor (NS)	c15 : Performance | c151 : Earnings | c152 : Earnings Projections |
Publisher Name (PUB)	c1521 : Analyst Comment/Recommendation | ccat :
Accession Number (AN)	Corporate/Industrial News International Associated Services Ltd.

Table 1: Extracts from Factiva News Data

3.5 Word Lists

The number of positive words and negative words of each article, as well as the tone of articles, is calculated from 5 different wordlists. Three of these will be domain-specific, whilst the remaining two will be general word lists. The domain-specific word lists are:

- $LM\ wordlist$
- $Henry\ 2006\ wordlist$
- $Henry\ 2008\ wordlist$

Additionally, the general word lists are also used are:

- $Diction\ 7\ wordlist$
- $LIWC\ wordlist$

The $LM\ wordlist$ was developed by Loughran and Mcdonald (2011) in their research of 10-K filings. The $Henry\ 2006$ and $Henry\ 2008\ wordlists$ were implemented by Henry (2006, 2008) in their study of earnings press releases. The $Diction\ 7\ wordlist$ is a general word list from the Diction 7.0 computer software (https://www.dictionsoftware.com/), whilst the $LIWC\ wordlist$ is from the computer software "Linguistic Inquiry and Word Count 2015" (Pennebaker *et al.* 2015).

3.6 Development of Tone Measures

In this study, sentiment is considered on a word and tone level and the corresponding variables are derived from the press releases and media articles by the previously mentioned wordlists, and are displayed in Table 1, where i, j, t, z, refers to article, company, day and wordlist, respectively.

Variable	Description
$W_{i,j,t,z}$	$Total\ number\ of\ words\ in\ article, i, about\ company, j, on\ day, t, using\ wordlist, z$
$W^{+}_{i,j,t,z}$	$Number\ of\ Positive\ words\ in\ article, i, about\ company, j, on\ day, t, using\ wordlist, z$
$W^{-}_{i,j,t,z}$	$Number\ of\ Negative\ words\ in\ article, i, about\ company, j, on\ day, t\ using\ wordlist, z$
$Tone_{i,j,t,z}$	$Tone\ of\ \ article, i, about\ company, j, on\ day, t, using\ wordlist, z$
I_t	$Total\ number\ of\ articles\ on\ day, t$

Table 2: Variable Definitions for Sentiment Analysis

Furthermore, the variables are formatted to collect a signal on an article level, rather than on a daily level. The tone of each article is calculated by subtracting the total amount of positive words, $W^{+}_{i,j,t,z}$, by the sum amount of negative words, $W^{-}_{i,j,t,z}$, and dividing that by the sum total number of positive and negative words $\left(W^{+}_{i,j,t,z} + W^{-}_{i,j,t,z}\right)$. Thus, $Tone_{i,j,t,z}$, is calculated by the following equation:

$$Tone_{i,j,t,z} = \frac{W^{+}_{i,j,t,z} - W^{-}_{i,j,t,z}}{W^{+}_{i,j,t,z} + W^{-}_{i,j,t,z}}$$

Equation 1

To format this on to a daily level, the sum of all the tones of each article published on day, t , will be divided by I_t. This daily tone will be calculated from the following equation:

$$Tone_{t,z} = \frac{\sum_{i=1}^{I_t}\left[Tone_{i,j,t,z}\right]}{I_t}$$

Equation 2

In addition, this study is also interested in investigating if there is a correlation throughout the tones calculated about each article by the wordlists. To do so, an average tone overall is calculated by adding together the tone of each individual wordlist and dividing it by the number of wordlists. This is shown in Equation 3.

Variable	Description
$Tone_{i,j,z,LM}$	$Tone\ of\ LM\ wordlist$
$Tone_{i,j,z,H6}$	$Tone\ of\ Henry\ 2006\ wordlist$
$Tone_{i,j,z,H8}$	$Tone\ of\ Henry\ 2008\ wordlist$
$Tone_{i,j,z,D7}$	$Tone\ of\ Diction\ 7\ wordlist$
$Tone_{i,j,z,LIWC}$	$Tone\ of\ LIWC\ wordlist$
$Tone_{i,j,z,AVG}$	$Average\ Tone$

Table 3: Different Tone Variables derived from each wordlist z, about article i, on company j on day

$$Tone_{i,j,z,AVG} = \frac{Tone_{i,j,z,LM} + Tone_{i,j,z,H6} + Tone_{i,j,z,H8} + Tone_{i,j,z,D7} + Tone_{i,j,z,LIWC}}{5}$$

Equation 3

These equations will be adapted to calculate the relevant numeric tone values of Media Articles and Press Releases, such that the tone will be calculated as:

$$MA\_Tone_{i,j,t,z} = \frac{W^{+}_{i,j,t,z} - W^{-}_{i,j,t,z}}{W^{+}_{i,j,t,z} + W^{-}_{i,j,t,z}}$$

Equation 4

And

$$PR\_Tone_{i,j,t,z} = \frac{W^{+}_{i,j,t,z} - W^{-}_{i,j,t,z}}{W^{+}_{i,j,t,z} + W^{-}_{i,j,t,z}}$$

Equation 5

The daily tones for Media Articles and Press Releases, will then be calculated as:

$$MA\_Tone_{t,z} = \frac{\sum_{i=1}^{I_t}\left[MA\_Tone_{i,j,t,z}\right]}{I_t}$$

Equation 6

And

$$PR\_Tone_{t,z} = \frac{\sum_{i=1}^{I_t}\left[PR\_Tone_{i,j,t,z}\right]}{I_t}$$

Equation 7

Finally, the average tones for Media Articles and Press Releases, will be calculated as:

$$MA\_Tone_{i,j,z,AVG} = \frac{MA\_Tone_{i,j,z,LM} + MA\_Tone_{i,j,z,H6} + MA\_Tone_{i,j,z,H8} + MA\_Tone_{i,j,z,D7} + MA\_Tone_{i,j,z,LIWC}}{5}$$

Equation 8

$$PR\_Tone_{i,j,z,AVG} = \frac{PR\_Tone_{i,j,z,LM} + PR\_Tone_{i,j,z,H6} + PR\_Tone_{i,j,z,H8} + PR\_Tone_{i,j,z,D7} + PR\_Tone_{i,j,z,LIWC}}{5}$$

Equation 9

For further graphical analysis, the articles are allocated in to various groups based on their average tone. The distinction for the two groups can be found below:

- Positive: When the tone is highly positive i.e. greater than 0.5.
- Negative: When the tone is negative i.e. less than 0.1.

The study of the correlation between the various wordlists tones, is undertaken to investigate if they are consistent in classifying a sources sentiment.

Chapter 4: The impact of Press Releases Tone Upon Media Articles Sentiment

4.1 Introduction

This chapter discusses the numerical sentiment data set used in the regression investigation and in addition, the research design, implementation and results obtained in the empirical analysis.

4.2 Dataset – News

As mentioned previously, this research obtained the data from the *Factiva* database. As completed by Ahern and Sosyura (2014) and Core *et al.* (2008), individual keys were assigned to each firm in order to obtain the articles. The dataset gathered contains a wide range of information over a time period initially between 2000 and 2014. It was then separated into two different categories, media articles and press releases. Similar to Bushee *et al.* (2010), *Factiva's* category of press release wires, were taken as the company disclosures.

4.3 Empirical Study - Sample Selection Financial Press Releases and Media Articles

The companies that are considered within the sentiment dataset, are from the Financial Times Share Exchange (FTSE) All-Share index, between the years 2000 and 2012. The focus is on companies of a considerable size due to the fact they will appear in the media on a regular basis, which will thus boost our analysis. The original dataset included a select few articles released in 2013 and 2014. However, as seen in Table 5, there are only 23 articles released in 2013, and 37 in 2014. In addition, all these articles are press releases, thus it was decided to exclude them from the dataset in further investigations discussed in more detail later on in the study. The overall sentiment dataset consists of a collection of press releases and media articles, 99,030 in total, relevant to 481 different companies. This information is displayed in the table below:

Total Number of Companies	481
Total Number of Publishers	216
Total Number of Releases	99030
Total Media Articles	54516
Total Press Releases	44514

Table 4: Summary of information available in the data set

Further examination of this sentiment dataset regarding their publication date, is seen in the tables below. Figure 1 is a bar chart detailing the Total number of articles released, and allocating them to the day of the week they were released. It is clearly evident from the graph that there is significant variation between the days in regards to the volume of articles being published. There is a visible increase over the week from a total of 11,899 on Mondays, which is 12.015% of the total number of releases, to 25,834 on Thursdays. This is an increase from 12.015% to 26.087%. Friday sees a deterioration in numbers as it falls down to 16,090, which is 16.248% of total contribution. However, it is also evident that there are very few publications on a Saturday or a Sunday, as only 2.286% and 1.359% of the total number of publications are released on these days. This may be due to the fact that the stock market is closed on such days.

Number of Total Releases, Media Articles and Press Releases per Day of the Week

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	Total
Total Releases	11899	20023	21574	25834	16090	2264	1346	99030
Media Articles	5491	10327	11749	13803	9654	2217	1275	54516
Press Releases	6408	9696	9825	12031	6436	47	71	44514

Number of Total Releases, Media Articles and Press Releases per Month of the Year

	January	February	March	April	May	June	July	August	September	October	November	December	Total
Total Releases	9771	8903	8625	8447	10413	6276	10343	7229	7307	7906	8972	4838	99030
Media Articles	6799	5166	4094	4306	5011	3006	5624	3929	3819	4970	5196	2596	54516
Press Releases	2972	3737	4531	4141	5402	3270	4719	3300	3488	2936	3776	2242	44514

Number of Total Releases, Media Articles and Press Releases per Year

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Total
Total Releases	4564	5649	6437	6825	6805	7500	7941	7759	8679	9354	10324	8578	8555	23	37	99030
Media Articles	2291	2963	3421	3465	3570	3826	3878	3904	4414	5636	6692	5190	5266	0	0	54516
Press Releases	2273	2686	3016	3360	3235	3674	4063	3855	4265	3718	3632	3388	3289	23	37	44514

Table 5: Volume of Total Releases, Media Articles and Press Releases by Calendar Period

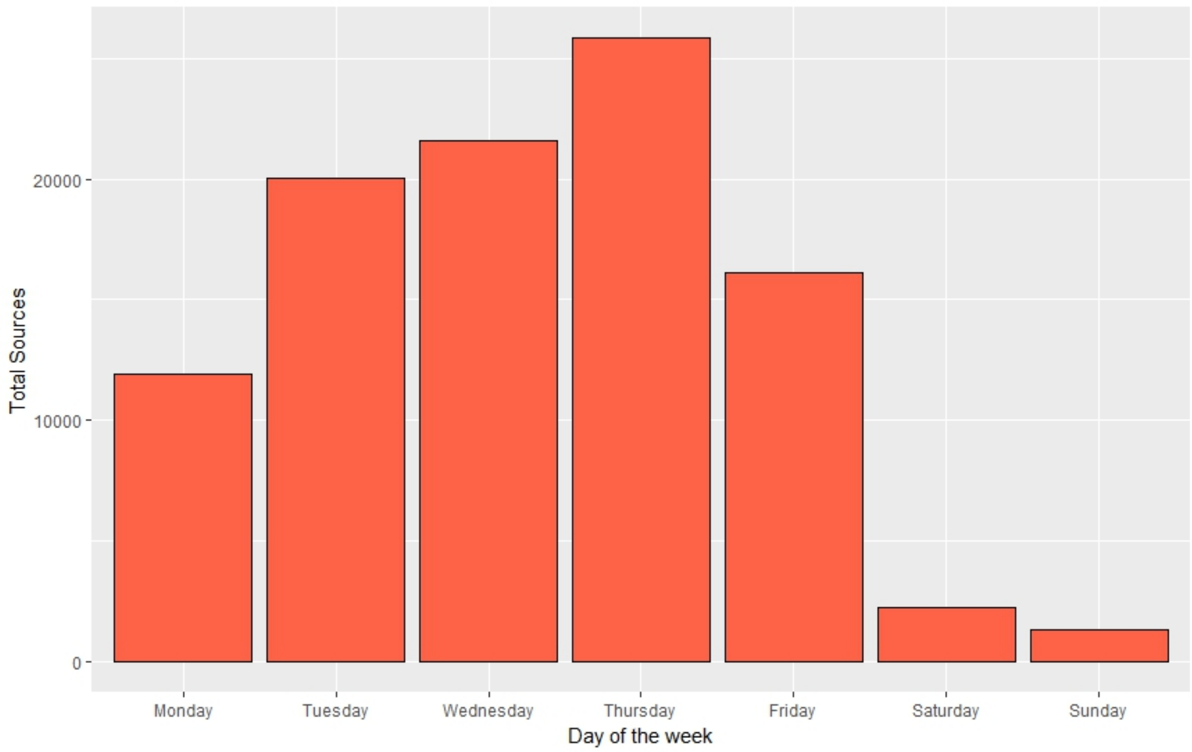


Figure 1: Number of News releases per day of the week over the time period: 2000 to 2014

In Appendix A, there are further graphs displaying the Total sources, Media articles and Press Releases against the calendar attributes, days of the week, months of the year and the year they were published. It can be seen in Appendix A, that the graphs for both Media Articles and Press Releases against days of the week, are similar in pattern to Table 5 above. Thursdays are evidently the day where the greatest number of articles are published with 25.319% of Media Articles being posted on this day, and 27.027% Press Releases.

In an attempt to gain a greater understanding of the information available, the average sentiment per month of both Media Articles and Press Releases is illustrated in the graph below, over the time period 2000 to 2012. This was achieved by adapting equations 6 and 7 to calculate the monthly sentiment, instead of the daily sentiment. Figure 2 shows that on average, the tone of Press Releases is higher than the tone of Media Articles, for most months of the year. This is shown in greater detail in Appendix C, where there are further tables displaying the average tone of Total sources, Media articles and Press Releases against the calendar attributes, days of the week, months of the year and the year they were published.

Additionally, months which see corporations issue their quarterly reports, accounts for 39.803% of the total Media Articles publications. These months are January, April, July and

October. However, these months only account for 33.1761% of the total number of company disclosures. This could be interpreted that the Media put greater focus on these months, as the information holds more weight. Furthermore, it is evident that over time, the volume of information releases has increased, where the year 2000 saw 4564 total releases and 2012 experienced 8555 in total. This is an increase of 87.45%

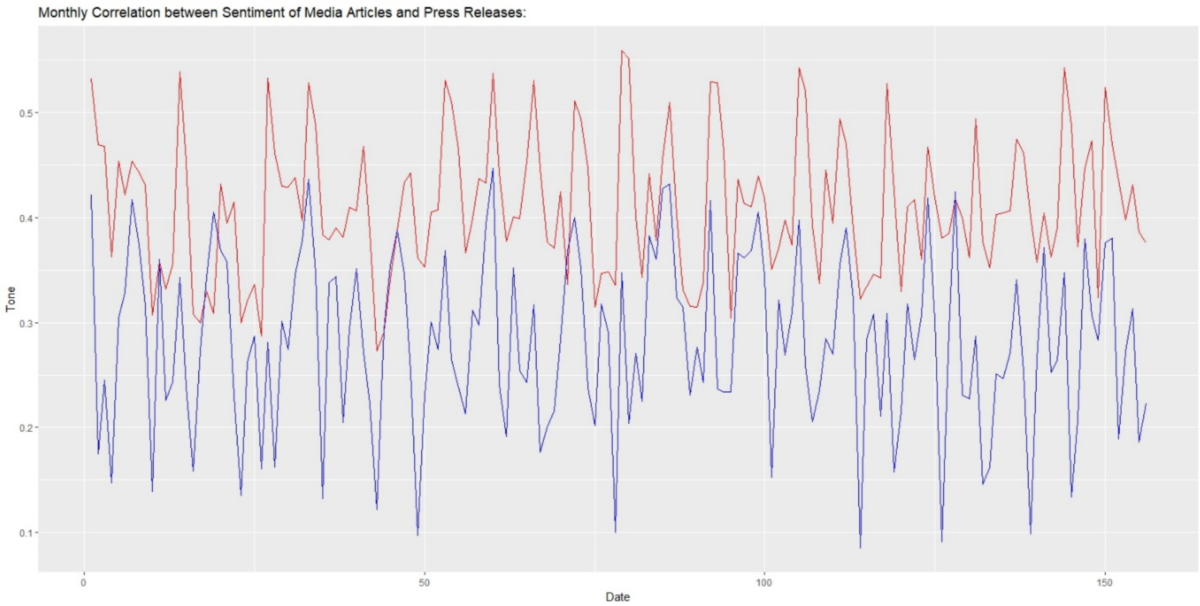


*Figure 2: Monthly Correlation between Sentiment of Media Articles and Press releases over the time period: 2000 to 2012**

* Note: (Media Articles = Blue, Press Releases = Red)

The tones generated by the wordlists were also investigated to study if there were positive correlations between the sentiment values they produce. As previously discussed, in order to fully comprehend the workings, the Average Tone was allocated into two groups, Positive and Negative. This was achieved by adapting Equation 3, to calculate the monthly sentiment, instead of the daily sentiment. Figure 3 shows the monthly values of the Average Tone, $Tone_{i,j,AVG}$, when it was Positive, and the corresponding average monthly values of the Tones from each of the five wordlists. Table 6 acts as a legend for Figure 3, to explain the graph. The values of $Tone_{H6}$ and $Tone_{H8}$ are seen to be quite similar, which would be expected as the same researchers implemented them (Henry 2006, 2008). This is also evident in Table 8, where it can be seen that the Pearson correlation value between PR_Tone_{H6} and PR_Tone_{H8} is 0.9949386, therefore highly correlated. The $Tone_{LM}$ calculated the most negative sentiment values of all the wordlists, never going over 0.6 in value, with its mean value equal to 0.4548, and its min value is equal to 0.2631. This is also consistent in Figure 37 in Appendix A.

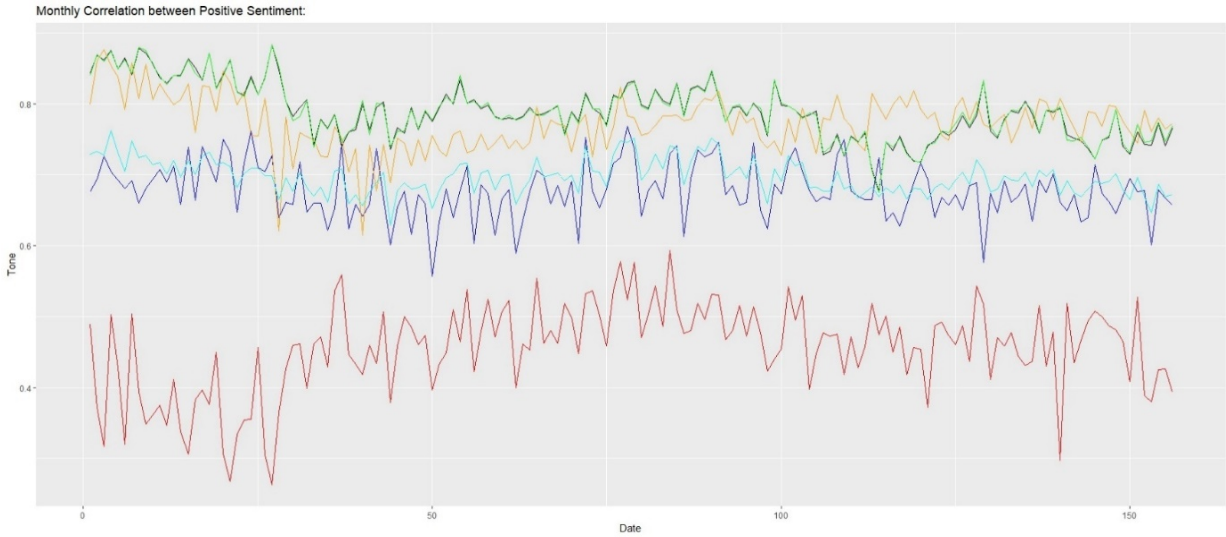


Figure 3: Monthly Correlation between Positive Sentiment

Variable	Description	Colours
$Tone_{LM}$	$Tone\ of\ LM\ wordlist$	Red
$Tone_{H6}$	$Tone\ of\ Henry\ 2006\ wordlist$	Black
$Tone_{H8}$	$Tone\ of\ Henry\ 2008\ wordlist$	Green
$Tone_{D7}$	$Tone\ of\ Diction\ 7\ wordlist$	Blue
$Tone_{LIWC}$	$Tone\ of\ LIWC\ wordlist$	Orange
$Tone_{AVG}$	$Average\ Tone$	Cyan

Table 6: Tone Variables and allocated colours for Figure 3

In order to compute the Pearson correlations, the dataset had to be first edited. Various days saw numerous articles published and the average tone was taken as the tone for that day. This was calculated from equations 8 and 9. To undertake the Pearson Correlations, our dataset was then transformed to only include days where both a Media Article and a Press Release was published. This resulted in a data set of 8867 observations. From Table 8, the word list which correlates most with PR_Tone_{AVG}, is the PR_Tone_{H6} with a value of 0.8663223. This is also consistent with the Media Articles sentiment, which recorded a value of 0.877218 between MA_Tone_{AVG} and MA_Tone_{H6}. The reasoning behind this could be due to two similar word lists being used to analyse the sentiment in the form of *Henry 2006* and *Henry 2008*. There is a surprising high value for the correlation between MA_Tone_{DIC} and MA_Tone_{LM}, 0.6442657, due to the fact the sentiment numerical values are computed from a specific and general dictionary. The least correlated tones seem to be PR_Tone_{LIWC} and PR_Tone_{H8}, where only a value of 0.3113311 was achieved.

	Obs	Mean	Std. Dev	Q1	Median	Q3
PR TONE LIWC	8867	0.6411332	0.2437396	0.5158803	0.6666667	0.8062102
PR TONE DIC	8867	0.4219601	0.3751059	0.1711837	0.452381	0.7048368
PR TONE H06	8867	0.4955396	0.3406913	0.3062302	0.5423729	0.7396017
PR TONE H08	8867	0.4983878	0.3398285	0.3157895	0.5467926	0.7407756
PR TONE LM	8867	0.1895171	0.4057262	-0.045455	0.1799569	0.4653168
PR TONE AVG	8867	0.4493076	0.2672501	0.2751661	0.4627751	0.6444777
MA TONE LIWC	8867	0.5249966	0.2998573	0.3333333	0.5555556	0.75
MA TONE DIC	8867	0.3335736	0.4329331	0.0508447	0.3684211	0.6666667
MA TONE H06	8867	0.4629109	0.3620128	0.2433147	0.5294118	0.7333333
MA TONE H08	8867	0.4659181	0.3614217	0.2476479	0.5317726	0.7380952
MA TONE LM	8867	0.0586697	0.4466052	-0.265082	0.0507937	0.3763553
MA TONE AVG	8867	0.3692138	0.3210729	0.1564467	0.3947475	0.6086672

Table 7: Descriptive Statistics for the Tone Measures of Media Articles and Press Releases

Table 7 above displays the summary statistics for each individual tone computed by the word lists. As mentioned previously, the data set contains 8867 observations due to it only including days where both a Press Release and a Media article was published. It is notable that the mean and median for all Press Release Tones are greater than the mean and median values for Media Articles. Thus, is another indication that Press Releases are more positive in sentiment values than the Media articles published about them.

	1	2	3	4	5	6	7	8	9	10	11	12
PR TONE LIWC	1											
PR TONE DIC	0.4760678	1										
PR TONE H06	0.3224307	0.4365847	1									
PR TONE H08	0.3113311	0.4277457	0.9949386	1								
PR TONE LM	0.35702	0.5081238	0.5828298	0.5790944	1							
PR TONE AVG	0.5858304	0.7419284	0.8663223	0.8606789	0.8072618	1						
MA TONE LIWC	0.3966372	0.2621608	0.3024837	0.30063	0.3114431	0.3940811	1					
MA TONE DIC	0.2838367	0.4079981	0.2929081	0.2923029	0.3428051	0.4194077	0.630408	1				
MA TONE H06	0.2875143	0.2449478	0.4344692	0.4338794	0.3253504	0.4411055	0.5634085	0.5430071	1			
MA TONE H08	0.2841296	0.2439178	0.434612	0.4354356	0.3224934	0.4397638	0.5516534	0.5354468	0.9959372	1		
MA TONE LM	0.287009	0.2845885	0.3307247	0.3285876	0.4239678	0.4288563	0.6710328	0.6442657	0.6320139	0.624277	1	
MA TONE AVG	0.3592773	0.3483177	0.4233161	0.4222645	0.4145373	0.5044954	0.7947162	0.8096576	0.877218	0.8708297	0.8603452	1

Table 8: Pearson Correlations of Tones generated by word lists for Media Articles and Press Releases

4.4 Model Design – Impact of Corporate Press releases on Media Articles

The model below is implemented to investigate if a notable correlation exists between the tone of company $j's$ information disclosures and the tone of the media article about company j, which follows this release, on day t.

$$MA\_Tone_{AVG} = \gamma_0 + \gamma_1 * PR\_Tone_{AVG} + \gamma_2 * PR\_Indicator_{AVG} + \gamma_3 * PR\_Tone_{AVG} * PR\_Indicator_{AVG} + \gamma_k * DAY + \gamma_l * MONTH + \gamma_m * YEAR + \varepsilon$$

Equation 10

The variables above in the model are described in the table below:

Variable	Description
MA_Tone_{AVG}	$Average\ Tone\ of\ Media\ Articles\ about\ company, j, on\ day, t.$
PR_Tone_{AVG}	$Average\ Tone\ of\ Press\ Releases\ about\ company, j, on\ day, t.$
DAY	$Fixed\ Effects\ per\ Weekday$
$MONTH$	$Fixed\ Effects\ per\ Month$
$YEAR$	$Fixed\ Effects\ per\ Year$
$COMPANY$	$Company\ Fixed\ Effects$
ε	*regression error term*

Table 9: Variable Definitions for Model investigating the correlation between Tone of Press Releases and Media Articles

The variable $PR\_Indicator_{AVG}$, is a dummy variable which focuses on a Press Releases median value by year quarter:

$$PR\_Indicator_{AVG} = \begin{cases} 1, & if\ PR\_Tone_{AVG}\ is\ > Median\ value\ (year - quarter) \\ 0, & if\ PR\_Tone_{AVG}\ is\ < Median\ value\ (year - quarter) \end{cases}$$

From the above model, the $PR\_Indicator_{AVG}$ variable was excluded in order to produce a notable coefficient γ_1. If the value of the coefficient, $\gamma_1$was equal 1, it would indicate that the publishers of media articles do include their own interpretation of the press releases, and publish the content without any change, or inclusion of different information. Therefore, if the coefficient γ_1 is significantly lower than 1, then it can be taken that the articles published by the Media take a less intense tone when reporting the same news events. The $PR\_Indicator_{AVG}$ is included to allow for the testing that if this moderation of tone is more evident for favourable new information releases.

4.5 Main Results

Dependent Variable MA_TONE_{AVG}	[1] Pooled OLS	[2] Firm Fixed Effects	[3] Pooled OLS	[4] Firm Fixed Effects
PR_TONE_{AVG}	0.606098 (55.014518)	0.578549 (51.248439)	0.522525 (25.207947)	0.505836 (24.589706)
$PR\_Indicator_{AVG}$			-0.06199 (-2.821449)	-0.058497 (-2.683552)
$PR\_TONE_{AVG} * PR\_Indicator_{AVG}$			0.158251 (4.28599)	0.145577 (4.043282)
$YEAR, MONTH, DAY$	Yes	Yes	Yes	Yes
R^2	0.2544	0.2285	0.2564	0.2403
Observations	8867	8867	8867	8867

Table 10: Regression Results using Pooled OLS and Firm Fixed Effects Models

The first part of this model solely concentrates on investigating the association between the numerical tone of corporations press disclosures, and the tone of the media articles which follow. In the first column of Table 10 above, the results of OLS Regressions between MA_Tone_{AVG} and PR_Tone_{AVG} are displayed, with the Fixed effects for $YEAR$, $MONTH$ and DAY. This reports that the estimated value of 0.606098 was achieved for the coefficient of PR_Tone_{AVG}, which is positive and the t-statistic equal to 55.014518. Additionally, all t-statistics in the Table above are recorded in brackets. These values indicate that there is a significant positive correlation between the tones of the two sources. The second column displays the results after a fixed effects regression was implemented to control characteristics which are unobservable on a corporation-level that are constant over time, where it is shown that the results do not change significantly. The third and fourth column presents the results after the values which are above or below the tones median value, are distinguished. This is achieved by incorporating the $PR\_Indicator_{AVG}$ variable into the model, and also the interaction between the variable and PR_Tone_{AVG}.

The results of the regressions show that the coefficients are significant with the values of 0.158251 and -0.06199 recorded for the coefficients of $PR\_Tone_{AVG} * PR\_Indicator_{AVG}$ and $PR\_Indicator$ respectively, whilst the t-statistics were recorded as 4.28599 and -2.821449. However, with the addition of the indicator, the PR_Tone_{AVG} coefficient equalled a predicted lower value of 0.522525 and a t-statistic of 25.207947. Nonetheless, it is still implied that media articles seem to consistently publish a less optimistic tone when reporting on corporations' press disclosures. These results do not significantly change in column 4, which display the results of the Fixed Effects Model.

The conclusion for this part of the study will be discussed in more detail in Chapter 7.

Chapter 5: GARCH and Aug-GARCH models

5.1 Introduction

This chapter discusses the methodology behind the GARCH-type models which will be implemented in this study to undertake the empirical analysis of News Volume impacting stocks volatility.

5.2 Description of the Models

Studies in recent years which focus on the stock returns' volatility, have provided results with substantial evidence for GARCH-type effects. Nonetheless, there is no rationale provided by these models as to how the flows of new information contribute to generating volatility. The work of Tauchen and Pitts' (1983), provided the Mixture of Distribution hypothesis (MDH) as an explanation. The theory of the MDH claims that at a given time period, the variance of log returns corresponds to the volume of information being reported. There are two assumptions which the MDH relies upon:

- A jointly independent distribution exists between the daily returns and the trading volumes (Harris 1987).
- The total amount of occurrences each day is stochastic.

The work of Lamoureux and Lastrapes (1990) provides support that there is a decrease in the predicted persistence when a stocks trading volumes are included in GARCH models. However, there is other work which provides evidence that some information to the returns process, is contributed by the trading volume.

An indicator that shocks to volatility remain over time, is if the GARCH(1,1) parameters α and β, are positive. The intensity of the persistence is reflected in the sum of these parameters, $\alpha + \beta$.

Variable	Description
δ_{it}	$the\ i^{th}\ \ intraday\ equilibrium\ price\ increment\ at\ day, t.$
n_t	$the\ number\ of\ information\ flows\ within\ day, t.$
σ^2	$Variance$
r_t	$Rate\ of\ Return$
μ_{t-1}	$Mean\ r_t\ conditional\ on\ past\ information$
Ω_t	$Conditional\ Variance$

Table 11: Variable Definitions

The work of Lamoureux and Lastrapes (1990) implies that:

$$\epsilon_t = \sum_{i=1}^{n_t} \delta_{it}$$

Equation 11

This is an indication that the stock returns follow a process of a linear combination of the movements of price. Furthermore, the work also considers:

$$\epsilon_t = r_t - \mu_{t-1}$$

Equation 12

Which is taken from the work of (Bollerslev 1986). Furthermore, δ_{it} is assumed to be Independent and identically distributed (IID) with mean zero and variance σ^2. Additionally, n_t is supposed to be quite large, which results in $\epsilon_t|n_t$ being asymptotically distributed as $N(0, \sigma^2 n_t)$. Ω_t is then defined as:

$$\Omega_t = \mathbb{E}(\epsilon_t|n_t)$$

Equation 13

If this mixture model is validated, then:

$$\Omega_t = \sigma^2 n_t$$

Equation 14

To ensure the argument being proposed is precise, it is assumed n_t, is serially correlated, such that:

$$n_t = a + b(L)n_{t-1} + \mu_t$$

Equation 15

Variable	Description
a	$constant$
$b(L)$	$a\ lag\ polynomial\ of\ order, q$
μ_t	$White\ noise$

Table 12: Variable Definitions

Substitution of the equations above, results in the following:

$$\Omega_t = \sigma^2 a + b(L)\Omega_{t-1} + \sigma^2 \mu_t$$

Equation 16

This above equation illustrates the dependence of the conditional variance, Ω_t upon the lagged variances, Ω_{t-1} and μ_t, the white noise. This equation was the focus of the work of Lamoureux and Lastrapes, (1990), to capture the persistence in conditional variance. However, the work of Shen *et al.* (2017) provides evidence that the association between this equation and the GARCH model is not uncomplicated. The main question being, "what proxy for n_t brings the best results?". One proxy for news arrivals which is commonly implemented is trading volumes. The theory behind this is the more news is published, the greater chance traders will classify the meaning of the information differently, therefore resulting in more traders being motivated to invest as their predictions for the future will vary. This proxy was implemented in the work of Lamoureux and Lastrapes (1990). It is assumed that ϵ_t is to be random, thus resulting in alternative distributions, with the variance depending on what information is available.

5.3 Model Design

There is a limited body of work which focuses on the investigating the influence on stock volatility. The work of Janssen (2004) examines the impact the frequency of information releases has on index volatility. This differs from the focus of this study, where the goal was to examine the effect on stock volatility. The theory covered in 5.1, lays the foundation for the study's models which come in the form of the following GARCH models:

Model 1: GARCH (1,1) model

A process, ϵ_t, will be considered where:

$$\epsilon_t = \sigma_t \mu_t, t \in \mathbb{Z}$$

Volatility, σ_t, is a nonnegative process, thus:

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

Equation 17

Model 2: Aug-GARCH(1,1) model – Volume $(\boldsymbol{V_t})$.

A process, ϵ_t, will be considered where:

$$\epsilon_t = \sigma_t \mu_t, t \in \mathbb{Z}$$

Volatility, σ_t, is a nonnegative process, thus:

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma V_t$$

Equation 18

Where

$$V_t = \frac{\upsilon_t}{\upsilon^*}$$

And the model parameters are ω, α, β and γ such that:

$$\omega > 0,$$

$$\alpha, \beta \geq 0,$$

$$\alpha + \beta < 1$$

Variable	Description
V_t	$Scaled\ daily\ trade\ volume\ of\ the\ stock\ at\ the\ day, t$
υ_t	$Daily\ trade\ volume\ of\ the\ stock\ at\ the\ day, t$
υ^*	$\max(\upsilon_t)$
σ_t	$volatility$

Table 13: Variable Definitions

Model 3: Aug-GARCH(1,1) model - Lagged Volume (V_{t-1}).

A process, ϵ_t, will be considered where:

$$\epsilon_t = \sigma_t \mu_t, t \in \mathbb{Z}$$

Volatility, σ_t, is a nonnegative process, thus:

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma V_{t-1}$$

Equation 19

And the model parameters are ω, α, β and γ such that:

$$\omega > 0,$$

$$\alpha, \beta \geq 0,$$

$$\alpha + \beta < 1$$

Variable	Description
V_{t-1}	$scaled\ trade\ volume\ for\ the\ company\ at\ day, t - 1$
r_t	$the\ log\ return\ of\ the\ company\ at\ the\ day, t$
r_t^*	$the\ log\ return\ of\ the\ FTSE100\ index\ at\ the\ day, t$
n_t	$number\ of\ all\ relevant\ news\ for\ the\ company\ released\ at\ the\ day, t$

Table 14: Variable Definitions

Model 4: Aug-GARCH(1,1) model - news intensity ($news_t$).

A process, ϵ_t, will be considered such that:

$$(\epsilon_t) = r_t(\theta_1 + \theta_2 r_t^*)$$

Equation 20

where

$$\epsilon_t = \sigma_t \mu_t, t \in \mathbb{Z}$$

Volatility, σ_t, is a nonnegative process, thus:

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma news_t$$

Equation 21

And the model parameters are $\omega, \alpha, \beta, \gamma, \theta_1$ and θ_2 such that:

$$\omega > 0,$$

$$\alpha, \beta \geq 0,$$

$$\alpha + \beta < 1$$

Model 5: Aug-GARCH(1,1) model - lagged news intensity ($\boldsymbol{news_{t-1}}$).

A process, ϵ_t, will be considered such that:

$$(\epsilon_t) = r_t(\theta_1 + \theta_2 r_t^*)$$

where

$$\epsilon_t = \sigma_t \mu_t, t \in \mathbb{Z}$$

Volatility, σ_t, is a nonnegative process, thus:

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma news_{t-1}$$

Equation 22

And the model parameters are $\omega, \alpha, \beta, \gamma, \theta_1$ and θ_2 such that:

$$\omega > 0,$$

$$\alpha, \beta \geq 0,$$

$$\alpha + \beta < 1$$

The GARCH(1,1) will be used to compare the results, to see if the inclusion of trading volume and intensity of news being published, improves or lessens the strength of the model. The second to fifth model, will be Aug-GARCH(1,1) models with trading volume and volume of news.

In later chapters of the study, it will be shown that Model 2: Aug-GARCH(1,1) model with V_t will compute results which indicate that GARCH effects are removed for some companies. However, Model 3: Aug-GARCH(1,1) with V_{t-1} does not remove these effects. Furthermore, Model 4: Aug-GARCH(1,1) with $news_t$, also does not remove GARCH effects. Finally, Model 5: Aug-GARCH(1,1) with $news_{t-1}$ does not remove GARCH effects.

5.4 Model Calibration – Quasi-Maximum likelihood estimations (QMLE)

Additionally, Quasi-Maximum likelihood estimates (QMLE) were implemented into this study in order to calibrate the models and to determine the unknown parameters. The Gaussian quasi-likelihood function is defined in the following equation:

$$L_n(\theta) = \prod_{t=1}^{m} \frac{1}{\sqrt{2\pi\tilde{\sigma}_t^2}} \exp\left(-\frac{\epsilon_t^2}{2\tilde{\sigma}_t^2}\right)$$

Equation 23

Furthermore, the log quasi-likelihood equation is defined as:

$$F_n(\theta) \coloneqq -\sum_{t=1}^{n} \left(\frac{\epsilon_t^2}{\tilde{\sigma}_t^2} log\tilde{\sigma}_t^2\right)$$

Equation 24

5.5 Model Testing - Monte Carlo Simulation – Loglikelihood Ratio

The likelihood ratio was used in this study to test the models against each other. This ratio has been discussed by Chan and Ling, (2006). It will be used to analyse the GARCH(1,1) model with the Aug-Garch(1,1) models. The loglikelihood ratio is defined by the following equation:

$$LLR = 2\left[LLF_{M_1}\left(\epsilon, \widetilde{\theta_0}\right) - LLF_{M_2}\left(\epsilon, \widetilde{\theta_1}\right)\right]$$

Equation 25

The variables are defined in the table below:

Variable	Description
M_1	Model 1: GARCH(1,1)
M_2	Model 2: Aug − GARCH(1,1) with volume
M_3	Model 3: Aug − GARCH(1,1) with lagged volume
M_4	Model 4: Aug − GARCH(1,1) with news
M_5	Model 5: Aug − GARCH(1,1) with lagged news
LLF_{M_i}	Loglikelihood Funtion for Model, i
LLR	Loglikelihood Ratio
ϵ_t	Random variable with mean and variance conditionally on information set, I_{t-1}

Table 15: Different Variables implemented into the Likelihood Ratio

Monte Carlo Simulation is used to approximate the asymptotic null distribution for Equation 25. M_1 is assumed to be the Null model, where $\widetilde{\theta_0}$ is the true parameter.

Chapter 6: The investigation of Aug-GARCH(1,1) Models with Volume of News Published

6.1 Dataset - Stock

The sample sentiment dataset covers a total of 481 companies on the FTSE100 index. However, the stock study focuses on 5 stocks traded on the London Stock Exchange, which were chosen due to having the greatest number of news publications. These companies are AstraZeneca, BP, GlaxoSmithKline, Tesco and Vodafone, with their daily closing prices, trading volume and news intensity being included in their dataset. The stock dataset was extracted from the financial dataset software, *Bloomberg.* This dataset covers the same time period as the sentiment's numerical dataset, January 1st, 2000 until December 31st, 2012, containing 3333 observations. The descriptive statistics for all 5 corporation's closing prices are displayed in the table below.

	From	To	Obs	Mean	Std. Dev	Min	Max
AstraZeneca	01/01/2000	31/12/2012	3333	2732.89	390.58	1748.00	3624.00
BP	01/01/2000	31/12/2012	3333	522.55	77.91	302.90	712.00
GlaxoSmithKline	01/01/2000	31/12/2012	3333	1381.65	252.39	987.00	2109.00
Tesco	01/01/2000	31/12/2012	3333	320.40	80.50	156.00	492.00
Vodafone	01/01/2000	31/12/2012	3333	175.68	59.10	95.92	475.44

Table 16: The Statistical values of the Chosen Stocks Daily Closing Prices

Furthermore, to gain a greater understanding of the stocks, the historical movements of each corporation were illustrated. These figures can be found in Appendix B whilst AstraZeneca's daily closing prices are depicted in Figure 7, alongside the historical movement of AstraZeneca's daily log returns. Table 17 reports the statistics for the stock indexes log returns, where it is shown that AstraZeneca's range in value from -12.6569 to 12.3569. Furthermore, it is reported that the value of means for the log returns is not statistically different from zero.

	From	To	Obs	Mean	Std. Dev	Min	Max	Skewness	Kurtosis
AstraZeneca	01/01/2000	31/12/2012	3332	2.27E-05	0.0173	-0.1266	0.1236	-0.174118	8.7412342
BP	01/01/2000	31/12/2012	3332	-2.07E-04	0.0173	-0.0976	0.1058	-0.078296	6.7529112
GlaxoSmithKline	01/01/2000	31/12/2012	3332	-5.18E-05	0.0155	-0.0909	0.1017	0.0469091	6.8085941
Tesco	01/01/2000	31/12/2012	3332	1.84E-04	0.0159	-0.1742	0.1221	-0.18557	11.25298
Vodafone	01/01/2000	31/12/2012	3332	-2.27E-04	0.0218	-0.1590	0.1195	-0.022038	7.3509478

Table 17: The Statistical values of the Chosen Stocks Daily Log Returns

From studying Figure 7, it is evident that over time the amplitude of the log returns for AstraZeneca fluctuates, thus showing signs of ARCH effects. The histogram of the AstraZeneca Log returns is illustrated in Figure 4, where the reported kurtosis value is 8.7412342, which is much greater than 3. Furthermore, the negative skewness indicates that the returns distribution tends to realize negative returns than those which are forecasted by the normal distribution. This is due to leptokurtic distributions.

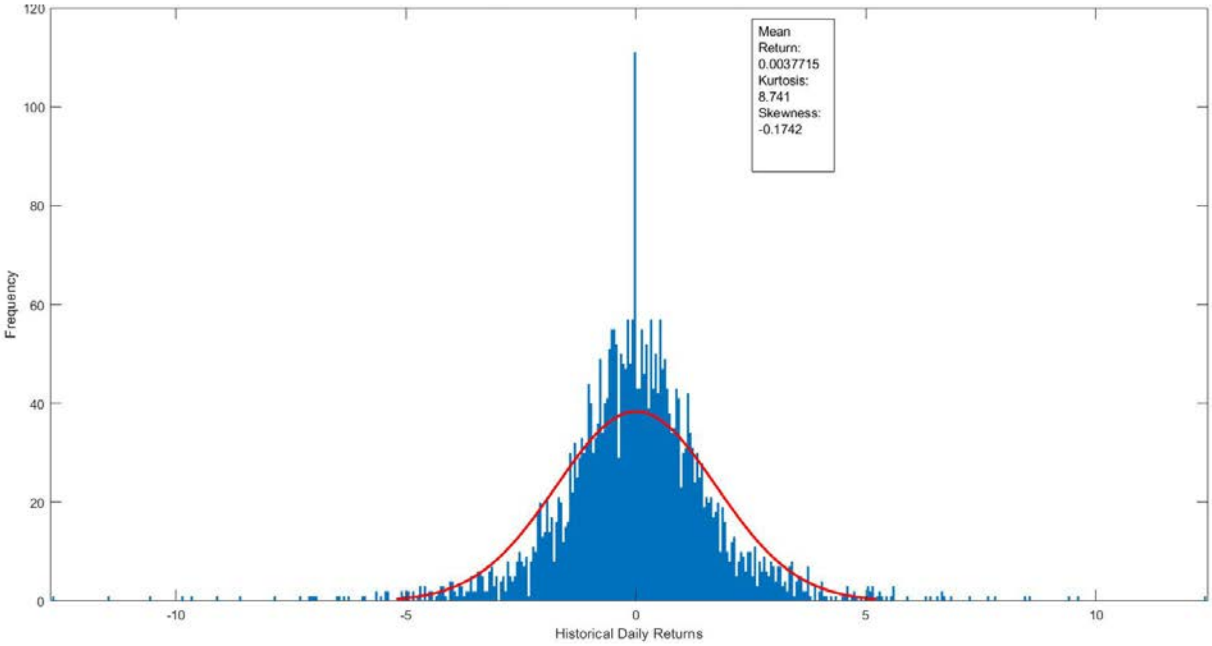


Figure 4: Histogram of AstraZeneca Log Returns

The autocorrelation for AstraZeneca's Daily returns and squared returns, are illustrated in Figure 5. The QQplot of AstraZeneca's daily returns and standard normal quantiles are shown in Figure 6. The daily historical movement of all the chosen corporations closing prices and returns, are shown in Appendix B, in addition to the Histograms of log returns, the autocorrelations of squared returns and returns, the QQ Plots of Daily Returns and Standard Normal Quantiles, for each company.

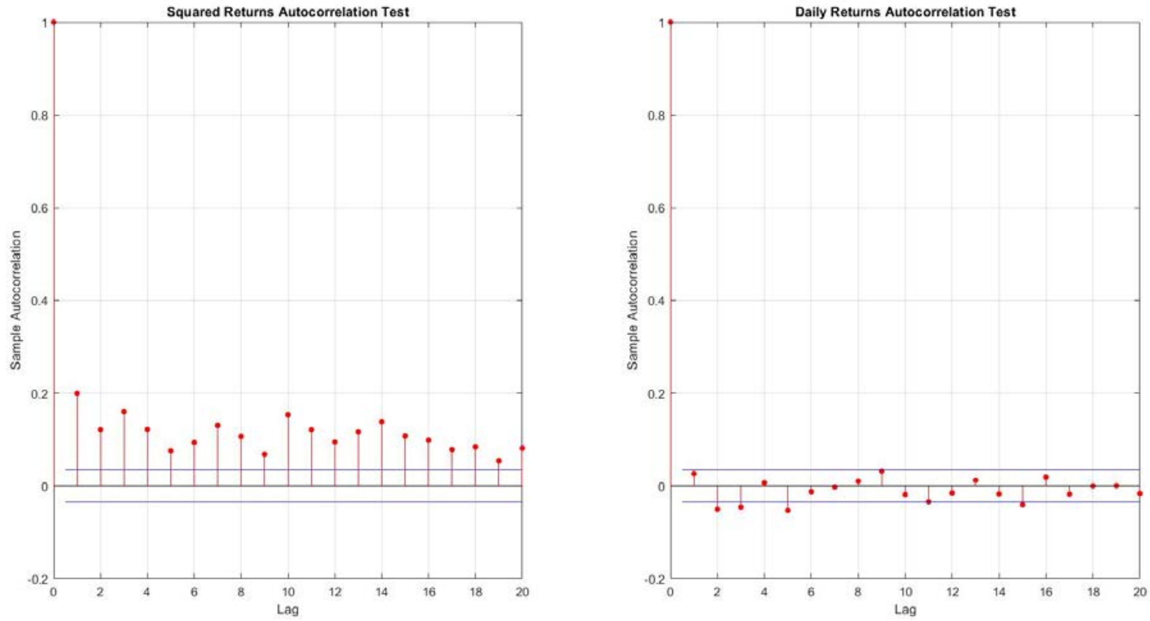


Figure 5: Autocorrelation for AstraZeneca's Squared Returns and Returns

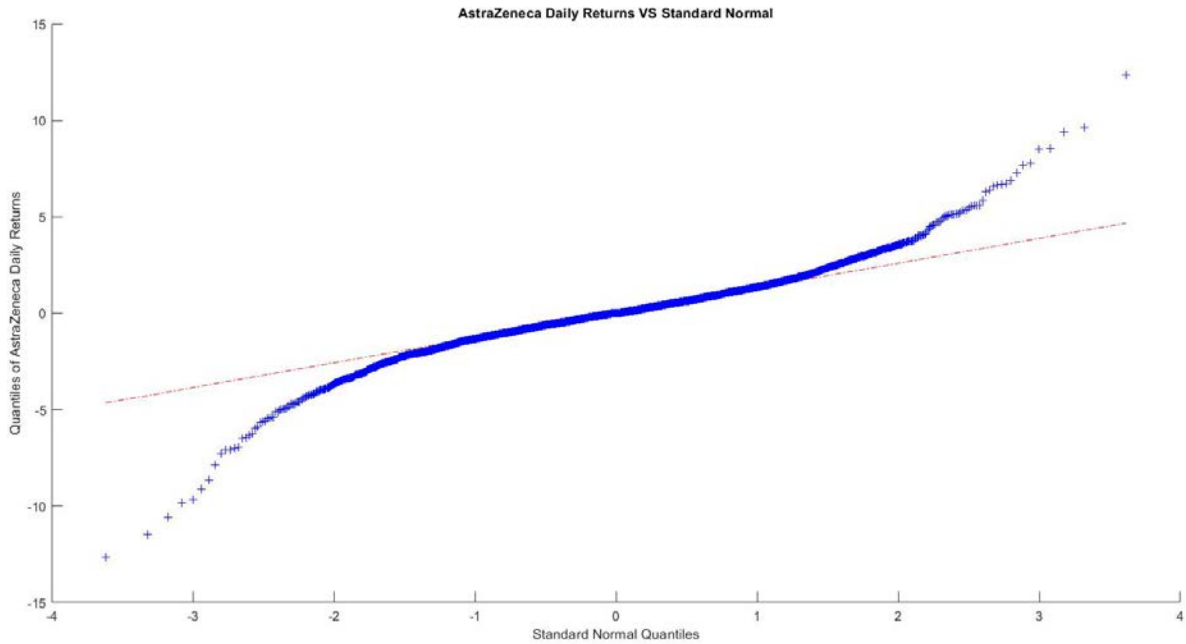


Figure 6: QQPlot of AstraZeneca Daily Returns and Standard Normal Quantiles

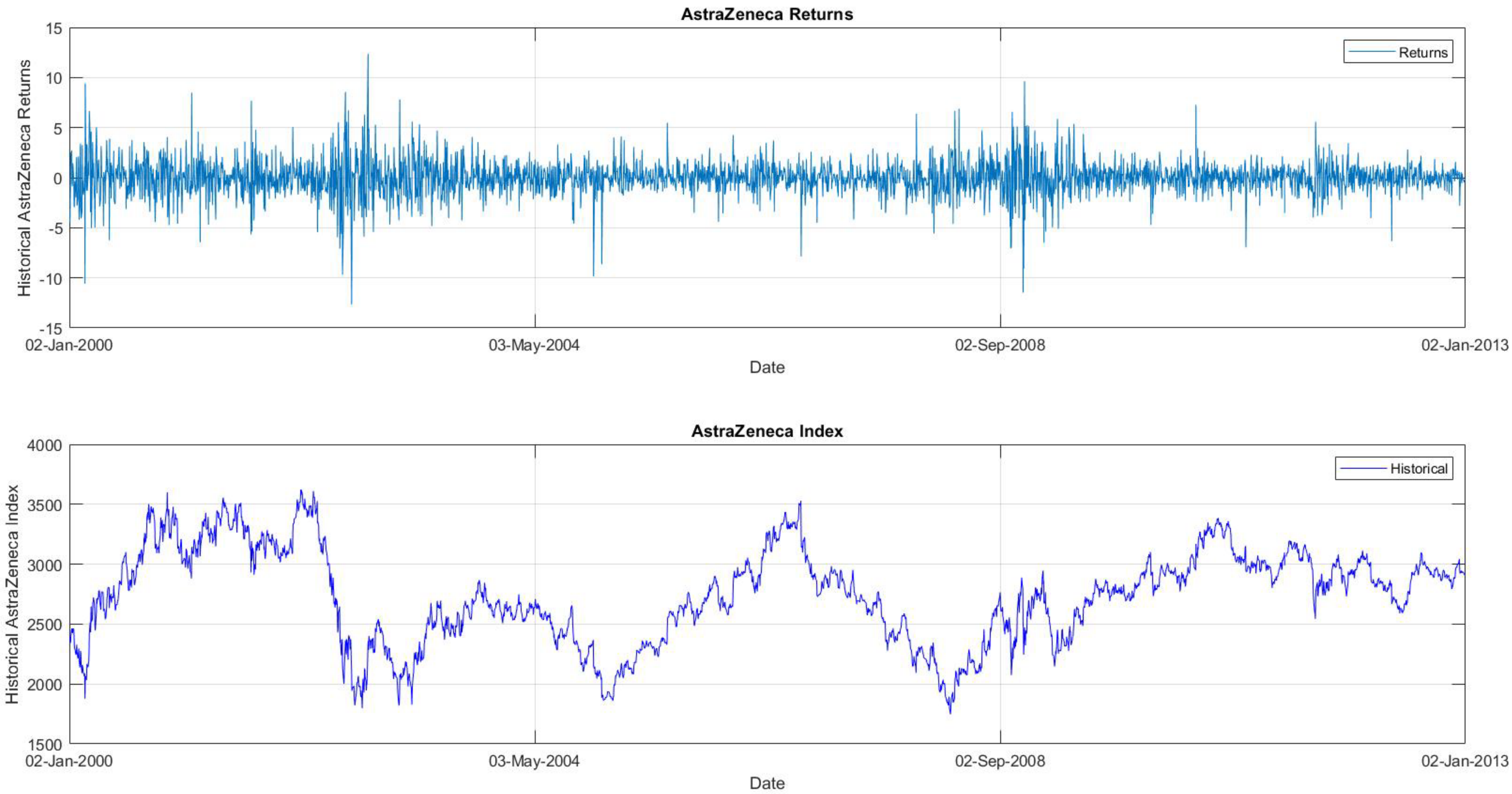


Figure 7: A comparison of AstraZeneca's Closing Daily Prices against its Historical Returns

6.2 Empirical Study

6.2.1 Model Parameters and Loglikelihood

The first model which was implemented was the GARCH (1,1), which was without the trading volume and the volume of news published, on the daily log returns. From Table 18 below, the estimates and coefficients for the GARCH model are shown. If $\alpha + \beta > 0.9$, then GARCH effects are evident.

Model 1: GARCH(1,1)

$$\sigma^2 = \omega + \alpha + \epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	$\boldsymbol{\alpha + \beta}$	**LLF**
AstraZeneca	0.0522627	0.9329392	0.9852019	12299.92
BP	0.0766021	0.9066643	0.9832665	12579.95
GlaxoSmithKline	0.1405174	0.77562	0.9161373	12721.89
Tesco	0.1108056	0.8480283	0.958834	12908.49
Vodafone	0.0542213	0.9419323	0.9961536	11997.75

Table 18: Estimates of Model 1 parameters

Comparing the results of Table 18 above and Table 19, it is clear that the inclusion of trading volume in Model 2, has a significant explanatory capability, in regards to the conditional volatility for 3 of the 5 chosen corporations, AstraZeneca, Tesco and Vodafone. On the other hand, BP or GlaxoSmithKline do not experience many changes in the $\alpha + \beta$ values, in comparison with the table above.

Model 2: Aug-GARCH(1,1) model – Trading Volume (V_t)

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma V_t$$

	α	β	γ	$\alpha+\beta$	LLF
AstraZeneca	0.1647884	0.1224641	0.0020008	0.2872525	12403.01
BP	0.0749647	0.9184444	1.633E-05	0.9934091	12573.68
GlaxoSmithKline	0.0854699	0.8828956	4.932E-05	0.9683656	12747.94
Tesco	0.1571347	0.016306	0.0016715	0.1734407	13110.38
Vodafone	0.0012445	1.53E-07	0.0019683	0.0012446	12590.26

Table 19: Estimates of Model 2 parameters

Model 3 estimates are reported in Table 20. The results in these tables do not indicate that there is evidence of vanishing GARCH effect with the addition of the lagged trading volume variable in regards to the conditional volatility of a companies' returns.

Model 3: Aug-GARCH(1,1) model - Lagged Volume (V_{t-1}).

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma V_{t-1}$$

	α	β	γ	$\alpha+\beta$	LLF
AstraZeneca	0.0532433	0.9281057	1.08E-05	0.9813491	12277.94
BP	0.0880602	0.8855392	3.09E-11	0.9735994	12561.20
GlaxoSmithKline	0.0263448	0.9697504	8.86E-06	0.9960952	12700.55
Tesco	0.0760137	0.8991356	2.88E-09	0.9751493	12892.44
Vodafone	0.0374864	0.9571255	1.06E-05	0.9946119	11989.68

Table 20: Estimates of Model 3 parameters

The estimates for Model 4 and 5, are presented in Table 21 and 22. The results in these tables indicate that there is no evidence of vanishing GARCH effect with the addition of the news volume variables in regards to the conditional volatility of a companies returns. After n_t is included in the equation, the sum of $\alpha+\beta$, are close in Tables 21 and 22 in comparison to Table 18.

Model 4: Aug-GARCH(1,1) model – News ($\boldsymbol{news_t}$)

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma news_t$$

	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	$\boldsymbol{\gamma}$	$\boldsymbol{\theta_1}$	$\boldsymbol{\theta_2}$	$\boldsymbol{\alpha+\beta}$	**LLF**
AstraZeneca	0.0374626	0.9540411	4.31E-06	-8.28E-06	0.7231868	0.9915	12848.43
BP	0.0373308	0.9565003	1.45E-06	-8.93E-05	0.9442929	0.99383	13502.15
GlaxoSmithKline	0.0244107	0.9705111	1.67E-06	2.07E-05	0.6375938	0.99492	13238.00
Tesco	0.0921541	0.9033019	3.77E-06	3.60E-04	0.606611	0.99546	13349.70
Vodafone	0.1093468	0.8765303	1.01E-05	1.90E-05	0.9257763	0.98588	12742.84

Table 21: Estimates of Model 4 parameters

Model 5: Aug-GARCH(1,1) model - Lagged News ($\boldsymbol{news_{t-1}}$).

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma news_{t-1}$$

	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	$\boldsymbol{\gamma}$	$\boldsymbol{\theta_1}$	$\boldsymbol{\theta_2}$	$\boldsymbol{\alpha+\beta}$	**LLF**
AstraZeneca	0.0231415	0.9738787	1.55E-06	-1.71E-05	7.26E-01	0.99702	12811.91
BP	0.0314187	0.9660654	7.02E-07	-1.52E-04	9.45E-01	0.99748	13487.05
GlaxoSmithKline	0.017716	0.9808884	4.40E-07	-1.46E-04	6.40E-01	0.9986	13229.62
Tesco	0.0851066	0.9137604	3.05E-06	3.49E-04	6.03E-01	0.99887	13307.88
Vodafone	0.085578	0.9107446	5.43E-06	1.34E-05	9.21E-01	0.99632	12682.63

Table 22: Estimates of Model 5 parameters

6.2.2 Model Simulations

Model 1: GARCH(1,1)

The volatility forecast from Model 1, for the company AstraZeneca is displayed below in Figure 8. The forecasts for the other corporations can be found in Appendix B.

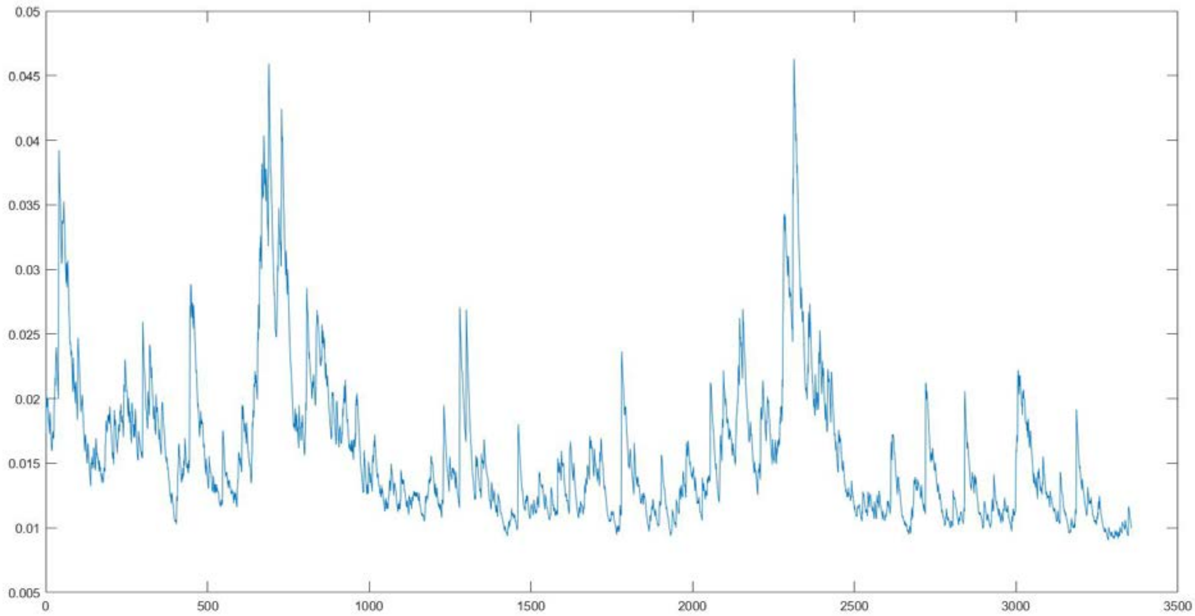


Figure 8: Volatility forecast of Model 1 for AstraZeneca returns

Model 2: Aug-GARCH(1,1) model – Volume (V_t).

The results reported in Table 23, are of the likelihood ratio test between Model 2 and Model 1. These were obtained after Monte Carlo simulations were completed, for a number of 1000 times.

	$M_2\ LFF$	$M_1\ LFF$	$2(M_2LFF - M_1LFF)$	$Null\ Hypothesis$
AstraZeneca	12403.01	12299.92	206.17	Rejected
BP	12573.68	12579.95	-12.53	Accepted
GlaxoSmithKline	12747.94	12721.89	52.11	Rejected
Tesco	13110.38	12908.49	403.78	Rejected
Vodafone	12590.26	11997.75	1185.02	Rejected

Table 23: The Likelihood Ratio test results between Model 1 and Model 2

The results recorded show that for the 5 chosen corporations, the null hypothesis is rejected for all except BP, with a confidence level of 1%.

The histograms of the parameters α and β for Model 2 after Monte Carlo simulations for AstraZeneca are illustrated in Figure 9 and Figure 10.

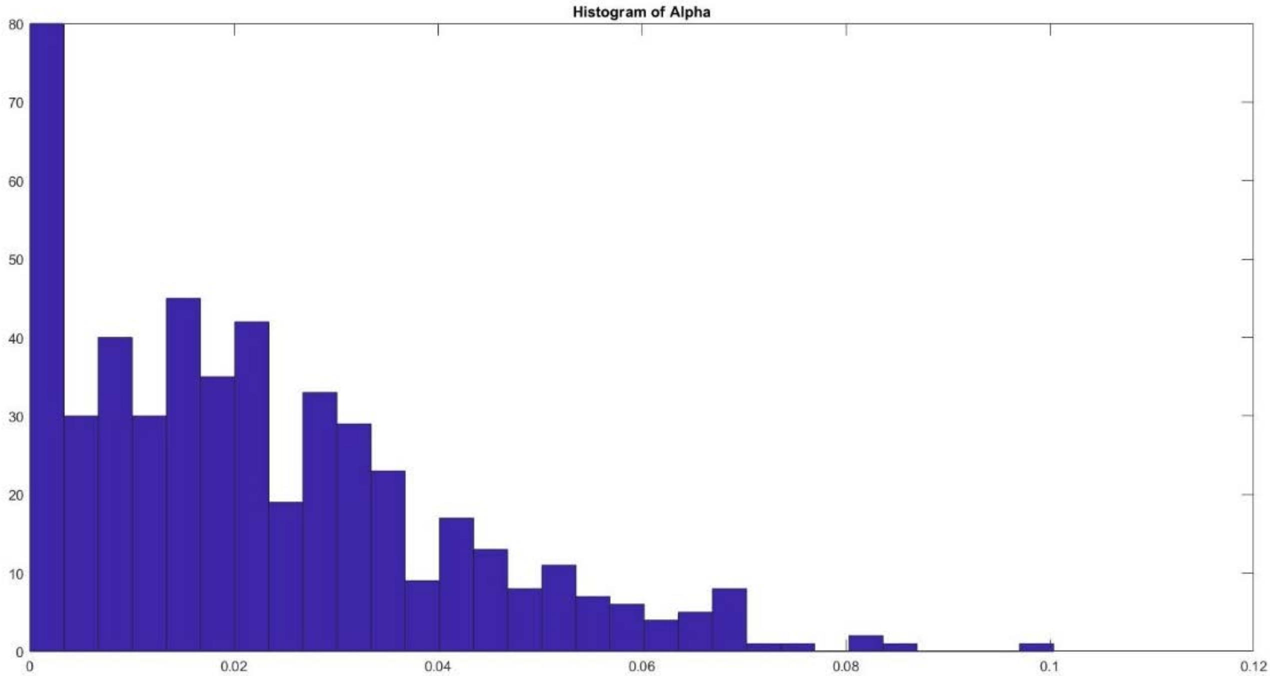


Figure 9: Model 2 - Histogram of α

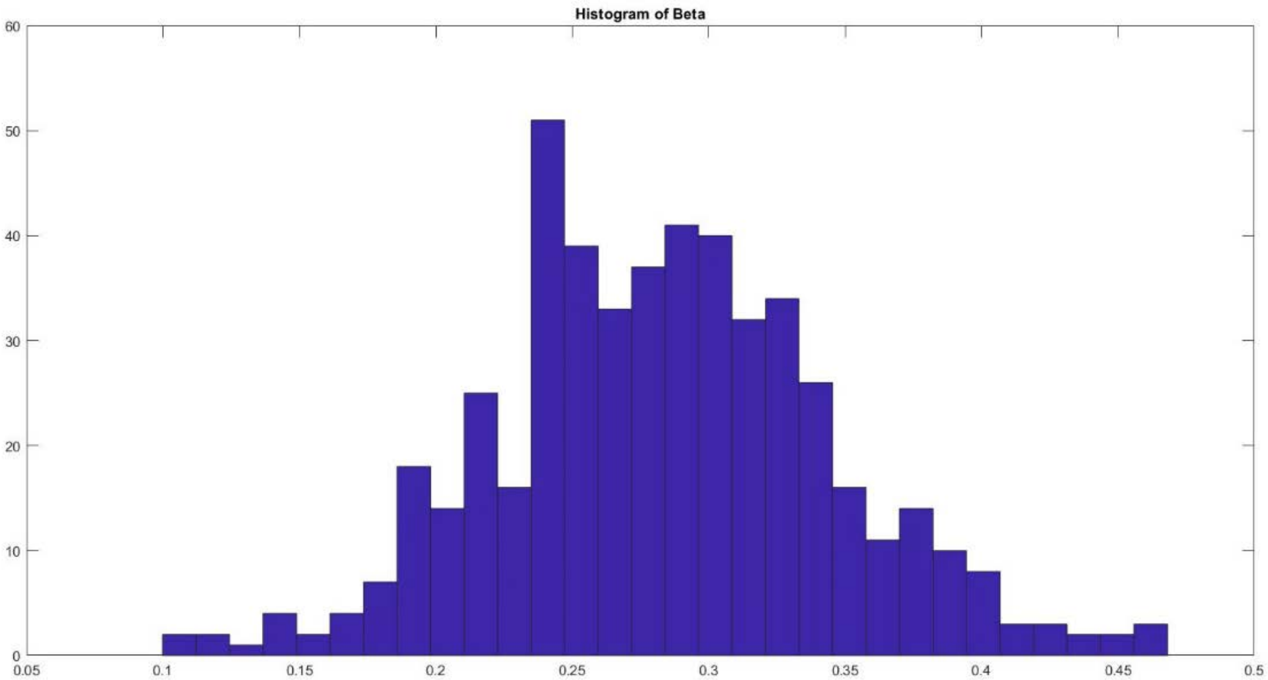


Figure 10: Model 2 - Histogram of β

The forecasted volatility for AstraZeneca, from the second model Aug-GARCH(1,1) with trading volume is shown in Figure 11 and Figure 12.

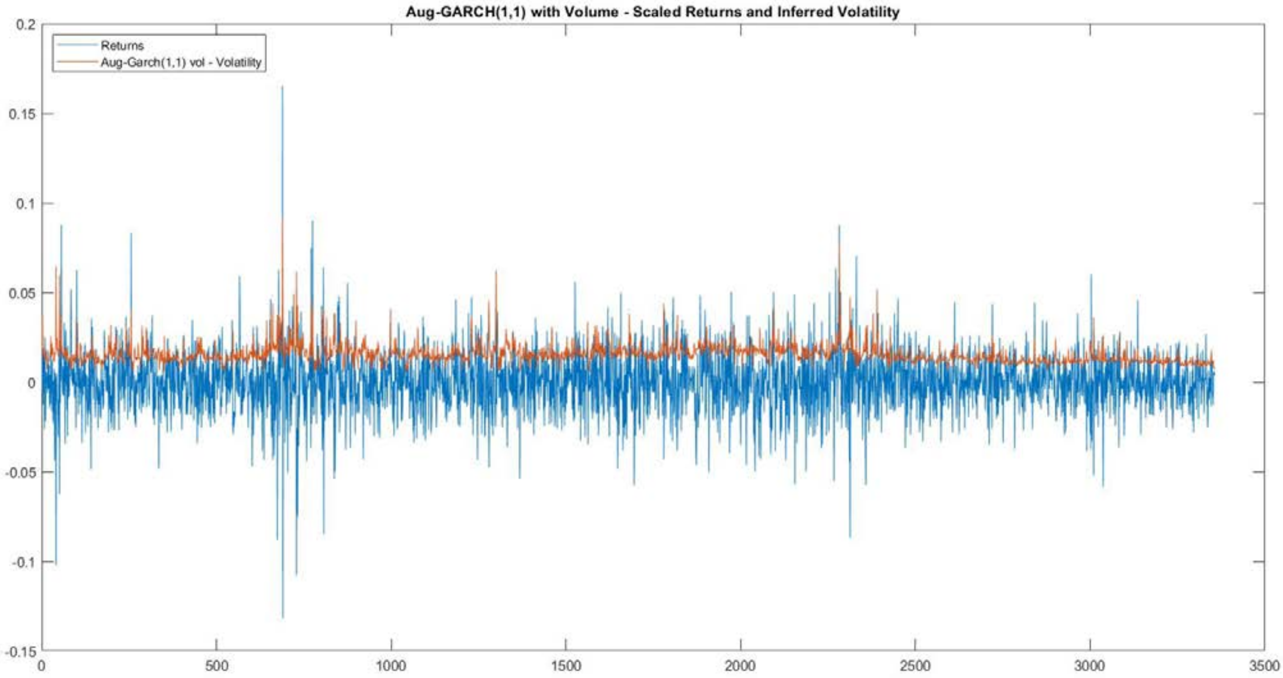


Figure 11: Inferred Volatility forecast and Scaled Returns of Aug-GARCH(1,1) with Volume model for AstraZeneca

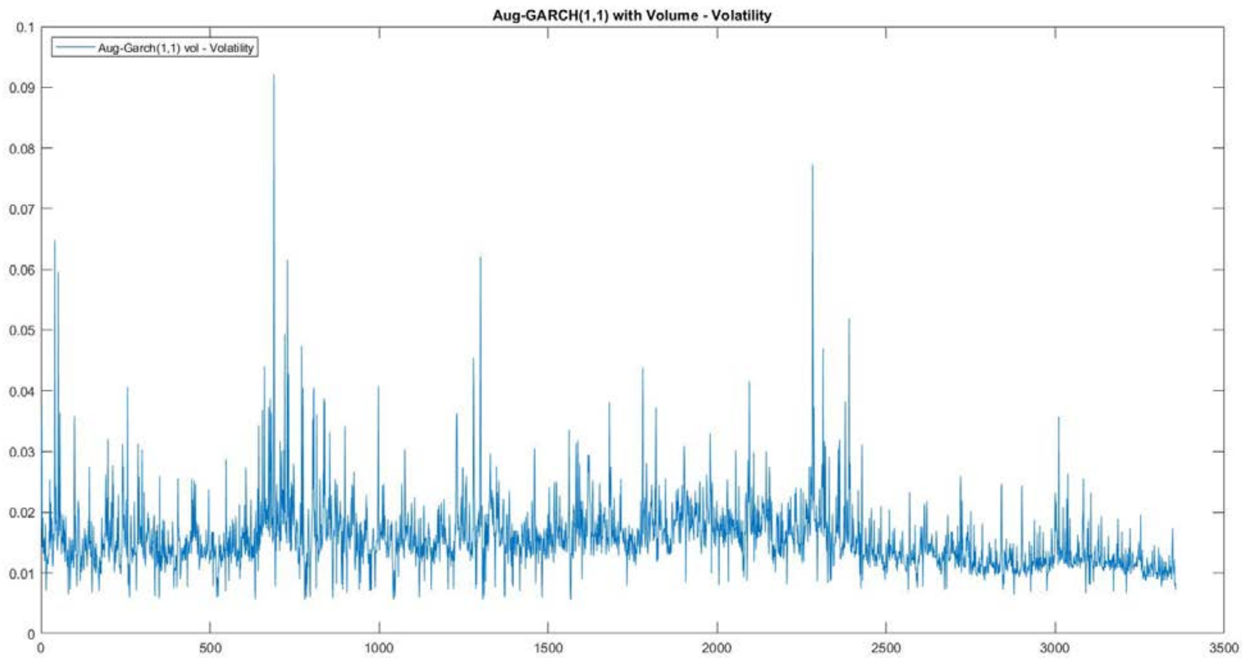


Figure 12: Volatility forecast of Aug-GARCH(1,1) with Volume model for AstraZeneca returns

Figure 13 is the histogram of the Loglikelihood ratio between Model 1 and Model 2 for AstraZeneca after the Monte Carlo simulations have been run 1000 times.

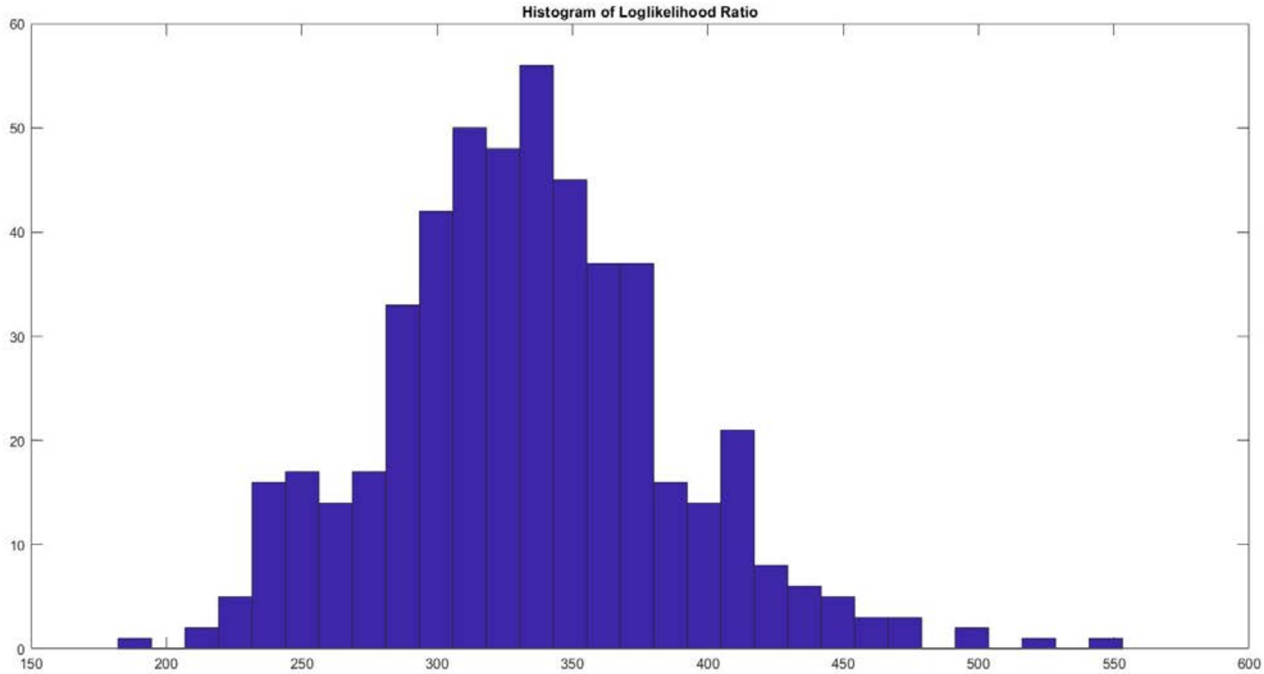


Figure 13: Histogram of Likelihood Ratio of Model 2

Model 3: Aug-GARCH(1,1) model - Lagged Volume (V_{t-1}).

The results reported in Table 24, are of the likelihood ratio test between Model 3 and Model 1, and also the significance of the parameters was studied. The number of Monte Carlo simulations to undertake was set at 1000. The results recorded show that for the 5 chosen corporations, the null hypothesis is accepted with a confidence level of 1%.

	M_3 LFF	M_1 LFF	$2(M_3LFF - M_1LFF)$	$Null\ Hypothesis$
AstraZeneca	12277.94	12299.92	-43.97	Accepted
BP	12561.20	12579.95	-37.50	Accepted
GlaxoSmithKline	12700.55	12721.89	-42.67	Accepted
Tesco	12892.44	12908.49	-32.10	Accepted
Vodafone	11989.68	11997.75	-16.14	Accepted

Table 24: The Likelihood Ratio test results between Model 1 and Model 3

The histograms of the parameters α and β for Model 3 after Monte Carlo simulations for AstraZeneca are illustrated in Figure 14 and Figure 15.

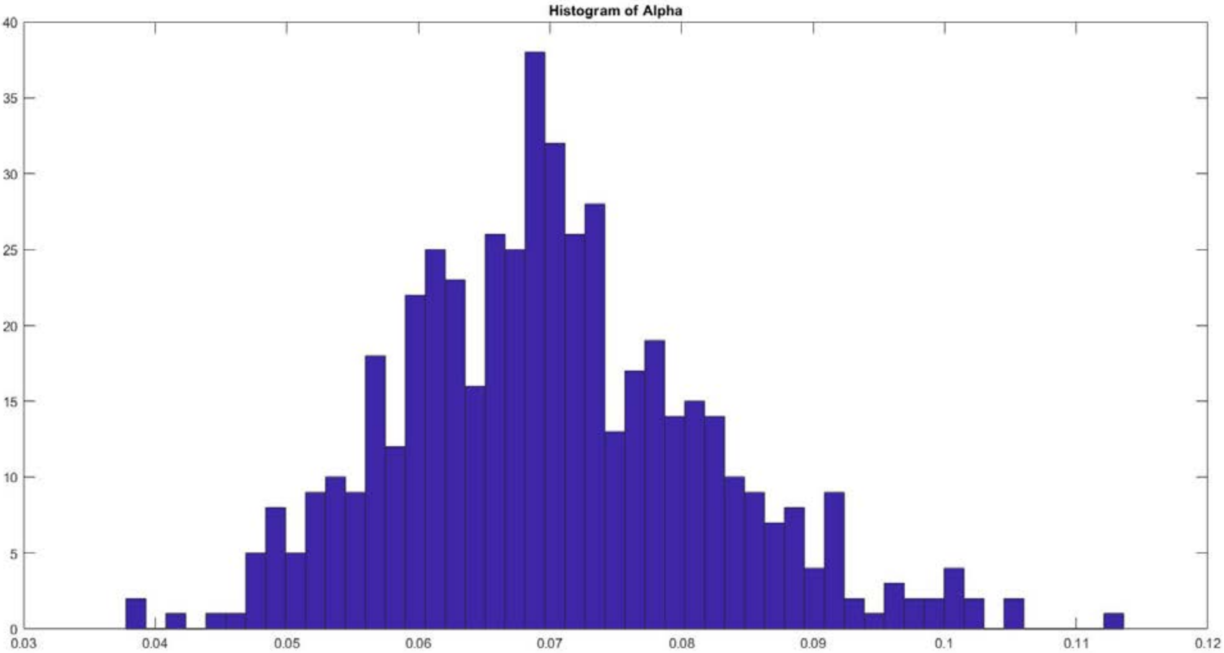


Figure 14: Model 3 - Histogram of α

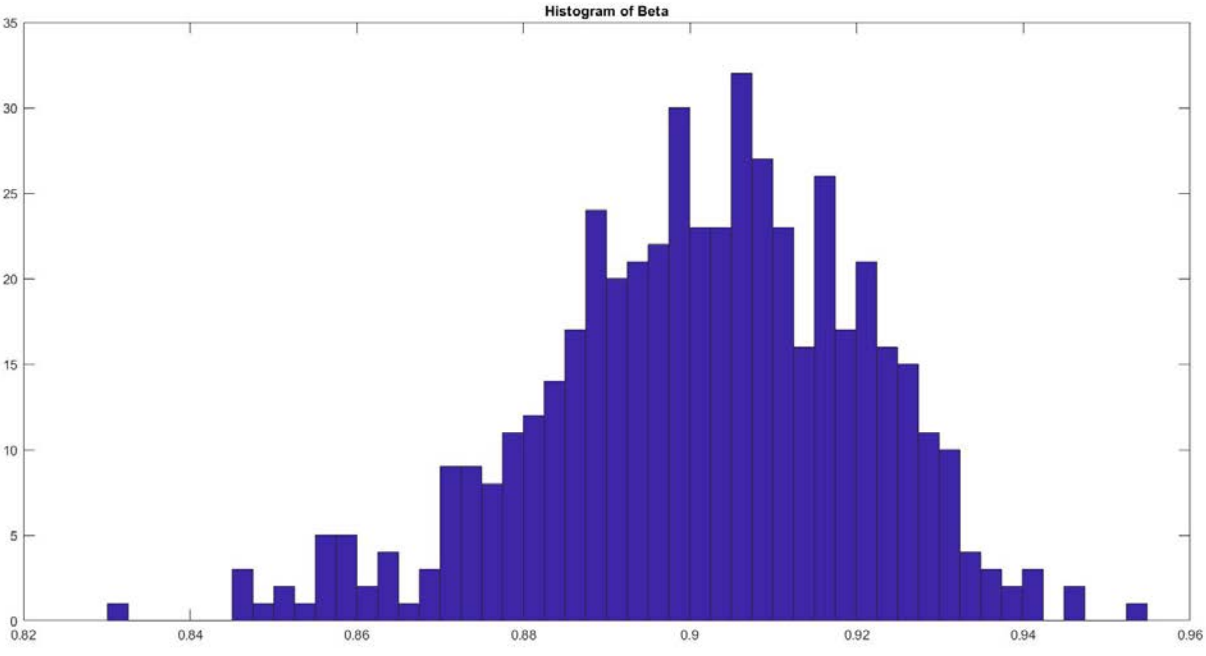


Figure 15: Model 3 - Histogram of β

The forecasted volatility for AstraZeneca, from the third model Aug-GARCH(1,1) with Lagged trading volume is shown in Figure 16 and Figure 17.

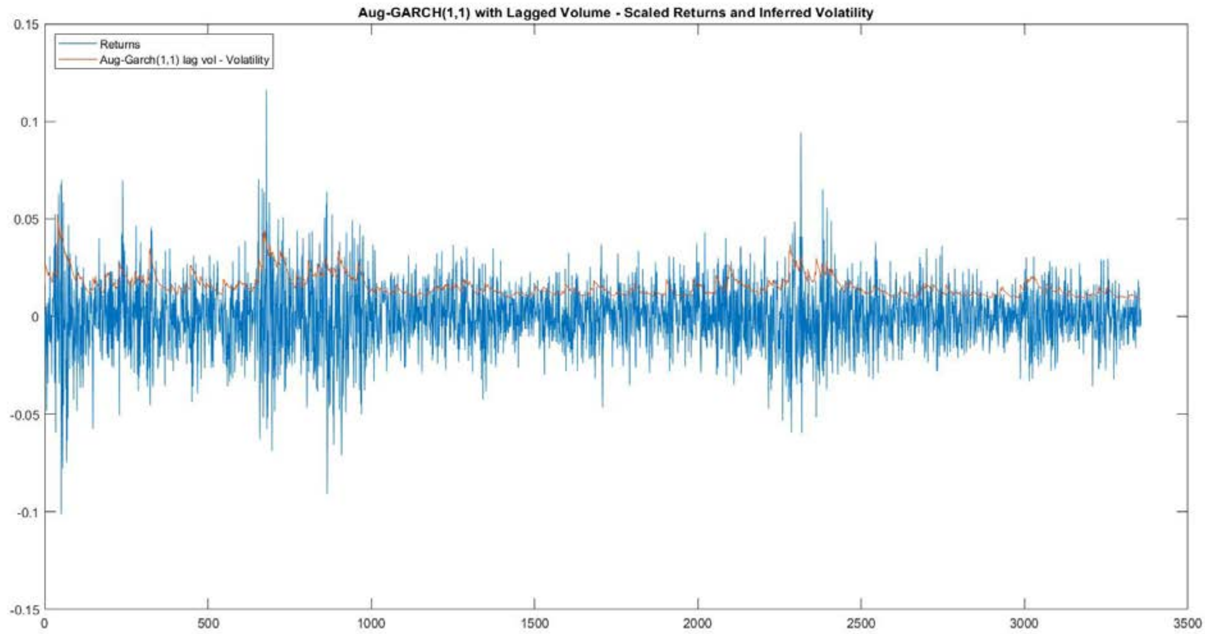


Figure 16: Inferred Volatility forecast and Scaled Returns of Model 3 for AstraZeneca

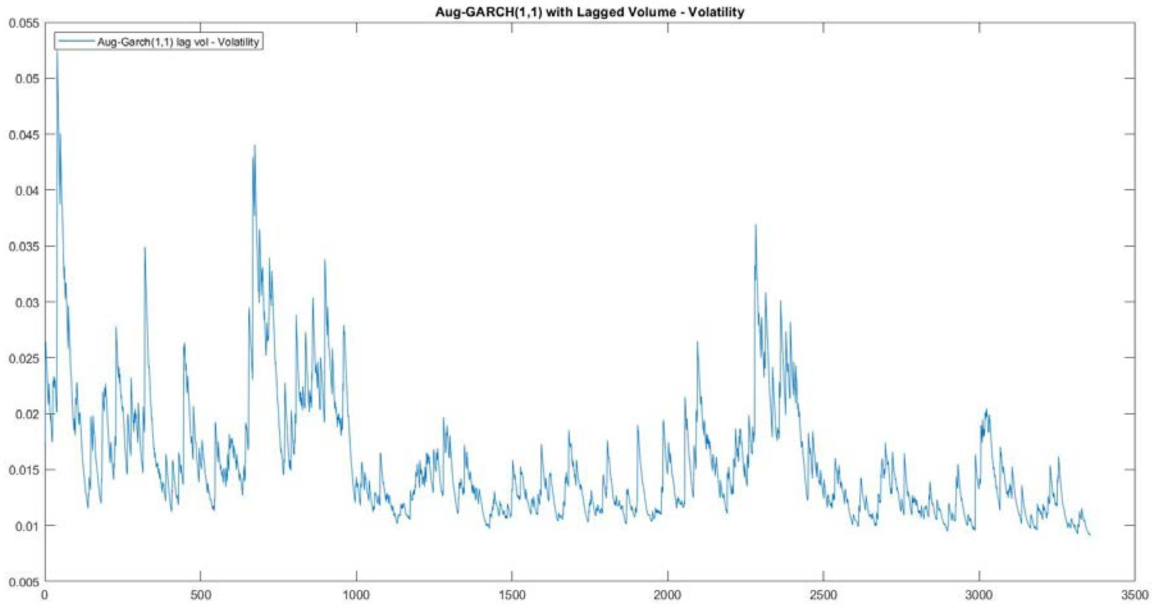


Figure 17: Volatility forecast of Model 3 for AstraZeneca returns

Figure 18 is the histogram of the Loglikelihood ratio between Model 1 and Model 3 for AstraZeneca after the Monte Carlo simulations have been run 1000 times.

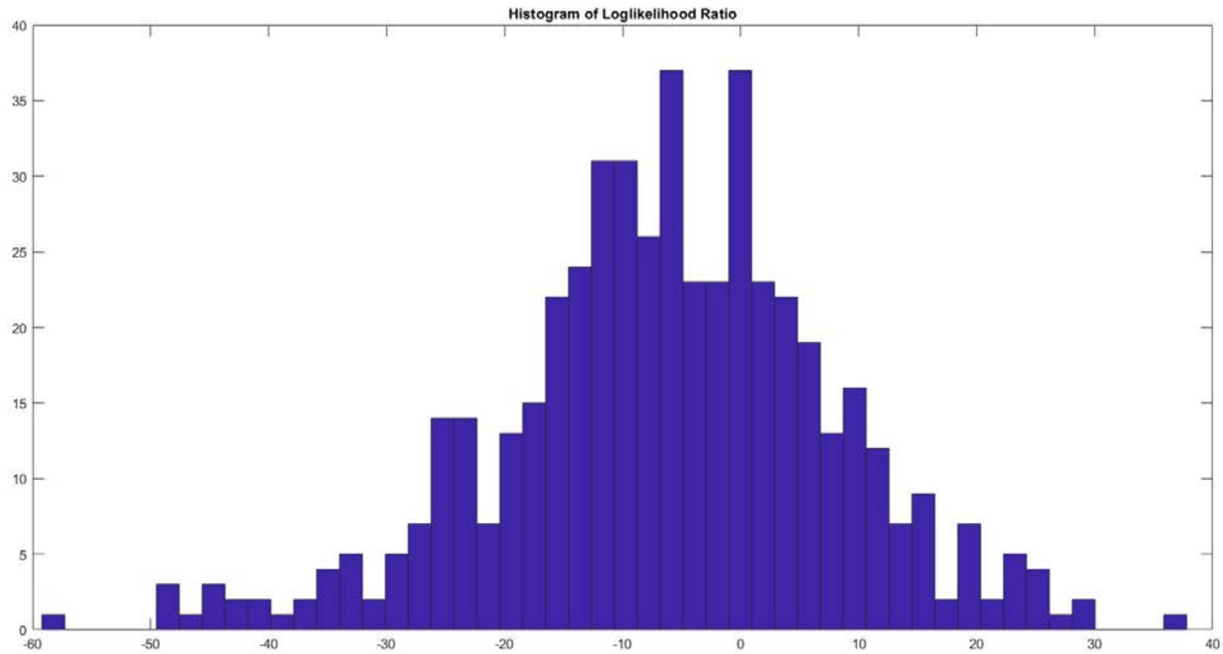


Figure 18: Histogram of Likelihood Ratio of Model 3

Model 4: Aug-GARCH(1,1) model – News ($news_t$).

The results reported in Table 25, are of the likelihood ratio test between Model 4 and Model 1. These were obtained after Monte Carlo simulations were completed for a number of 1000 times. The results recorded show that for the 5 chosen corporations, the null hypothesis is rejected with a confidence level of 1%.

	$M_4\,LFF$	$M_1\,LFF$	$2(M_4LFF - M_1LFF)$	$Null\ Hypothesis$
AstraZeneca	12848.43	12299.92	1097.02	Rejected
BP	13502.15	12579.95	1844.40	Rejected
GlaxoSmithKline	13238.00	12721.89	1032.23	Rejected
Tesco	13349.70	12908.49	882.42	Rejected
Vodafone	12742.84	11997.75	1490.17	Rejected

Table 25: The Likelihood Ratio test results between Model 1 and Model 4

The histograms of the parameters α and β for Model 4 after Monte Carlo simulations for AstraZeneca are illustrated in Figure 19 and Figure 20.

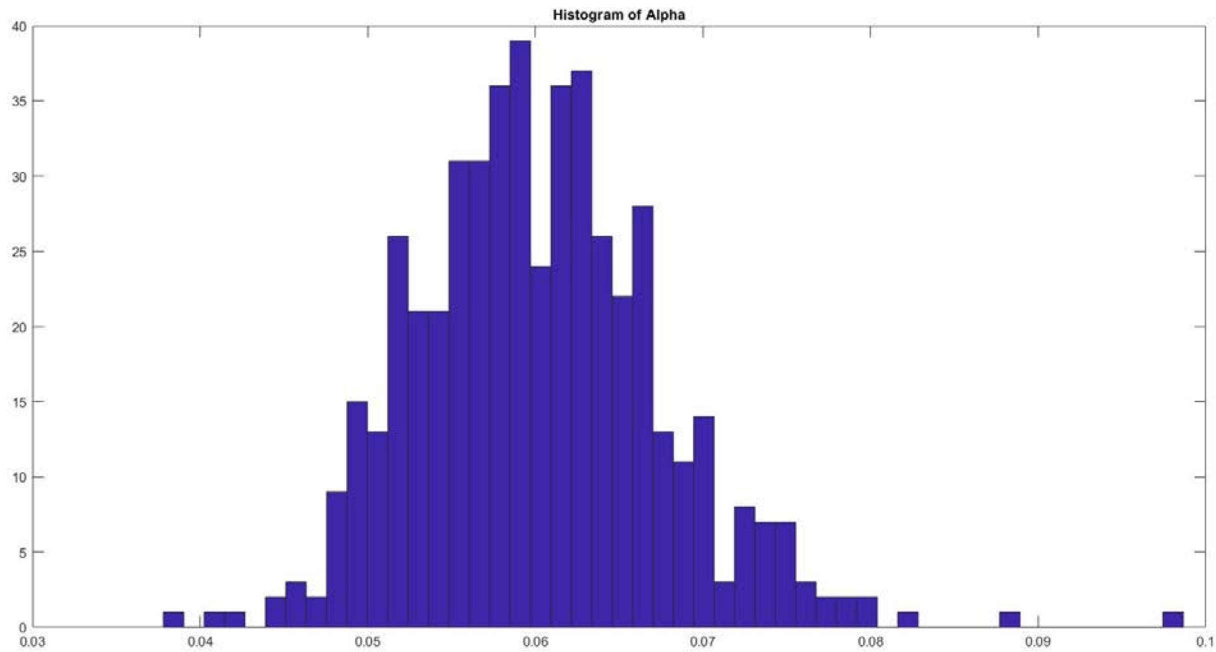


Figure 19: Model 4 - Histogram of α

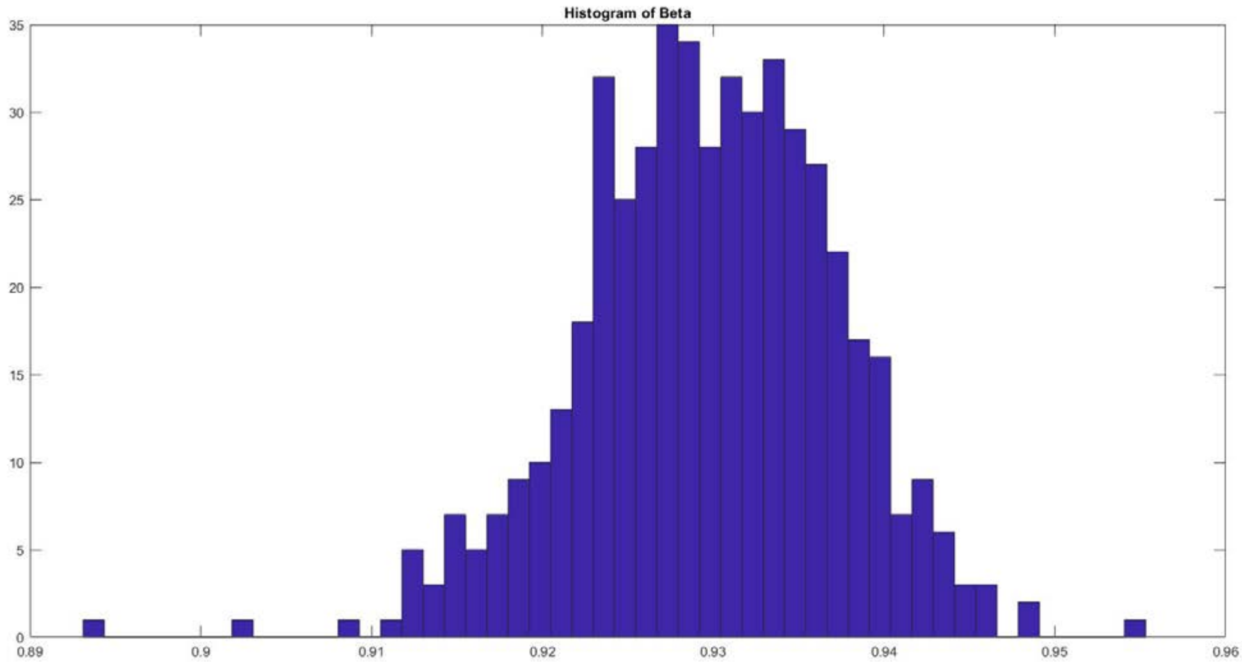


Figure 20: Model 4 - Histogram of β

The forecasted volatility for AstraZeneca, from the fourth model Aug-GARCH(1,1) with the volume of news published is shown in Figure 21 and Figure 22.

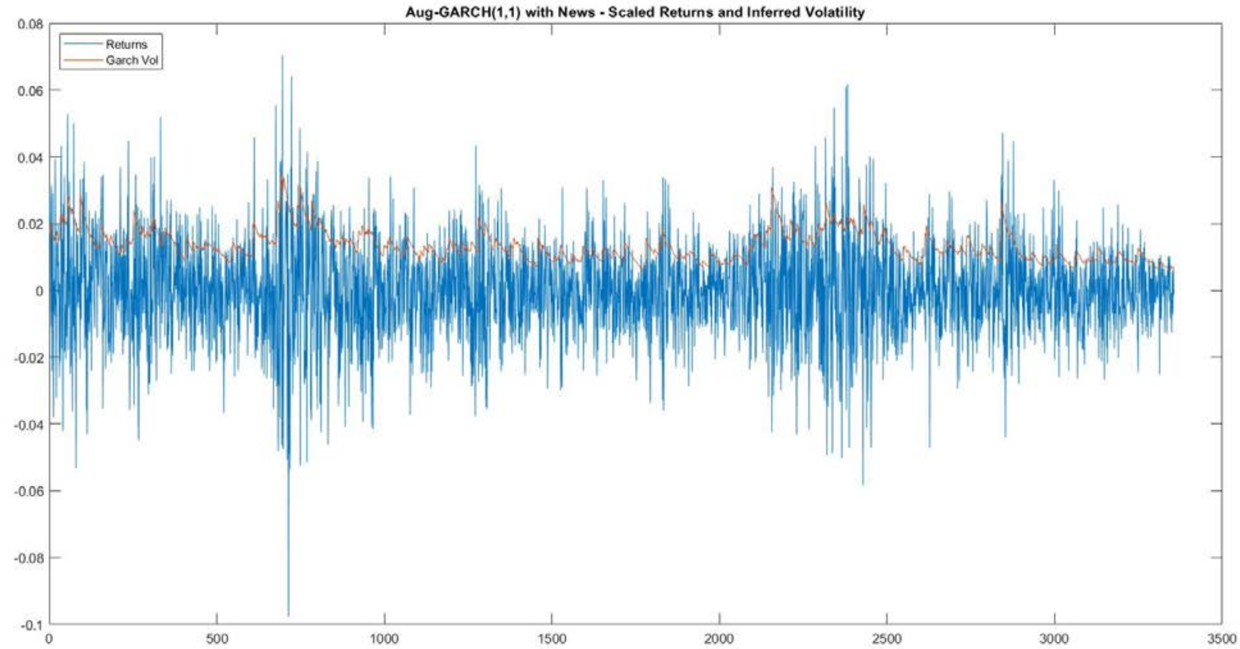


Figure 21: Inferred Volatility forecast and Scaled Returns of Aug-GARCH(1,1) with News model for AstraZeneca

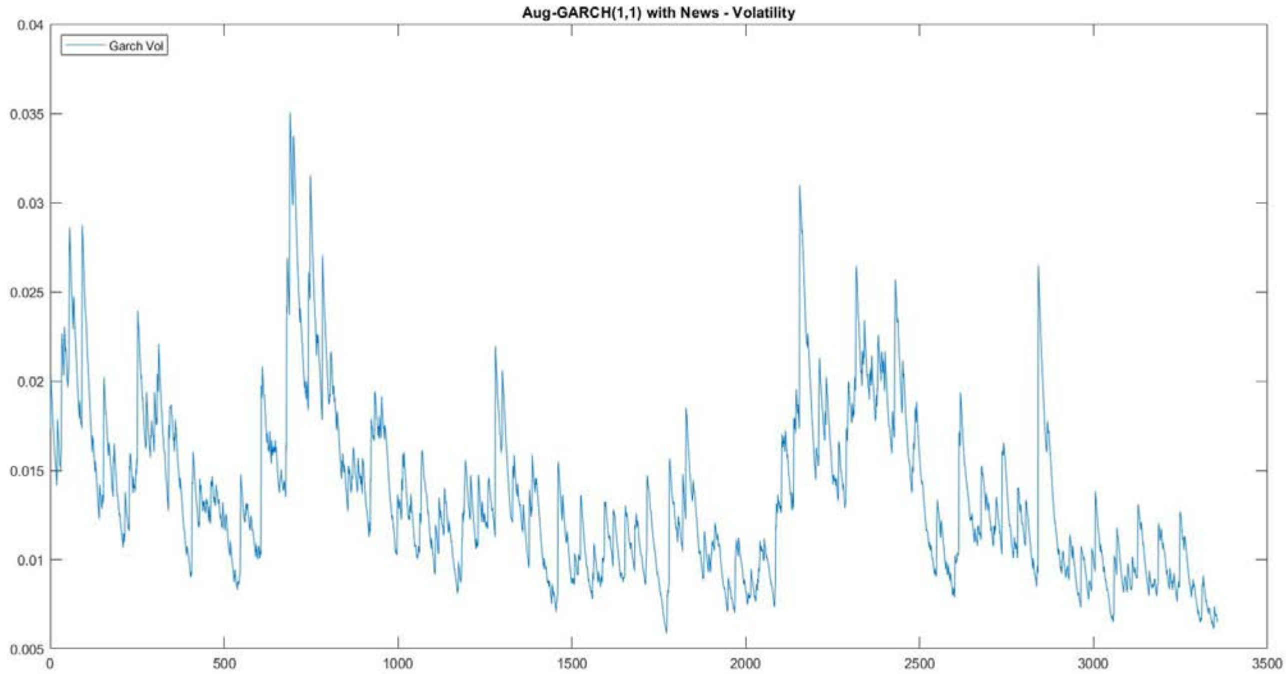


Figure 22: Volatility forecast of Model 4 for AstraZeneca returns

Figure 23 is the histogram of the Loglikelihood ratio between Model 1 and Model 4 for AstraZeneca after the Monte Carlo simulations have been run 1000 times.

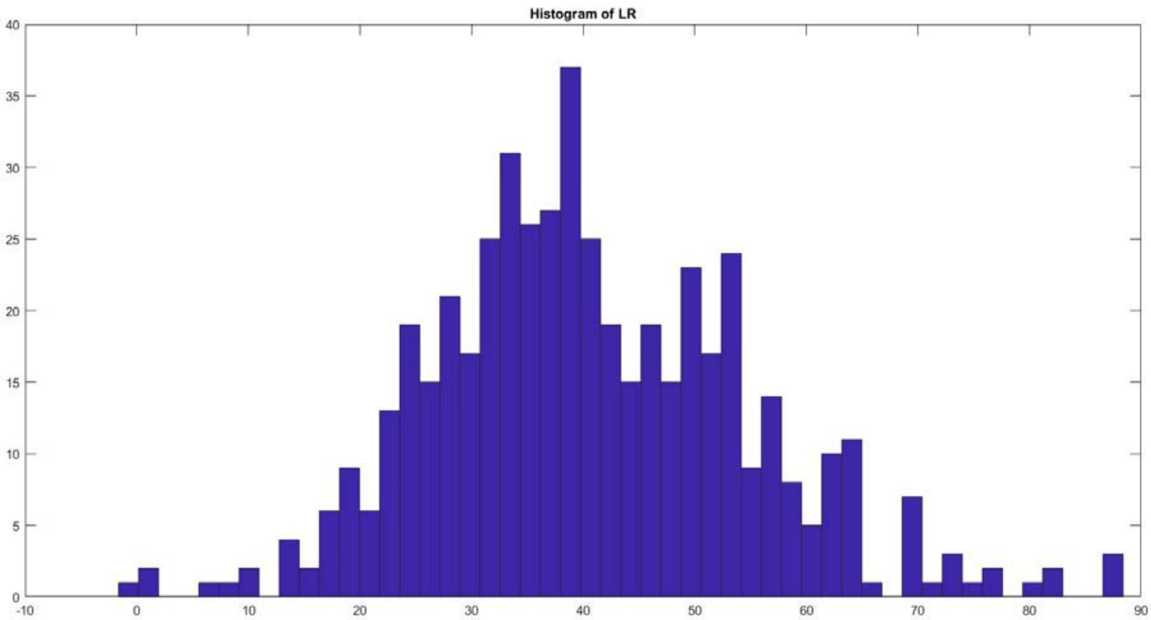


Figure 23: Histogram of Likelihood Ratio of Model 4

Model 5: Aug-GARCH(1,1) model - Lagged News ($news_{t-1}$).

The results reported in Table 26, are of the likelihood ratio test between Model 5 and Model 1. These were obtained after Monte Carlo simulations were completed for a number of 1000 times. The results recorded show that for the 5 chosen corporations, the null hypothesis is rejected with a confidence level of 1%.

	$M_5\ LFF$	$M_1\ LFF$	$2(M_5LFF - M_1LFF)$	$Null\ Hypothesis$
AstraZeneca	12811.91	12299.92	1023.98	Rejected
BP	13487.05	12579.95	1814.20	Rejected
GlaxoSmithKline	13229.62	12721.89	1015.46	Rejected
Tesco	13307.88	12908.49	798.77	Rejected
Vodafone	12682.63	11997.75	1369.75	Rejected

Table 26: The Likelihood Ratio test results between Model 1 and Model 5

The histograms of the parameters α and β for Model 5 after Monte Carlo simulations for AstraZeneca are illustrated in Figure 24 and Figure 25.

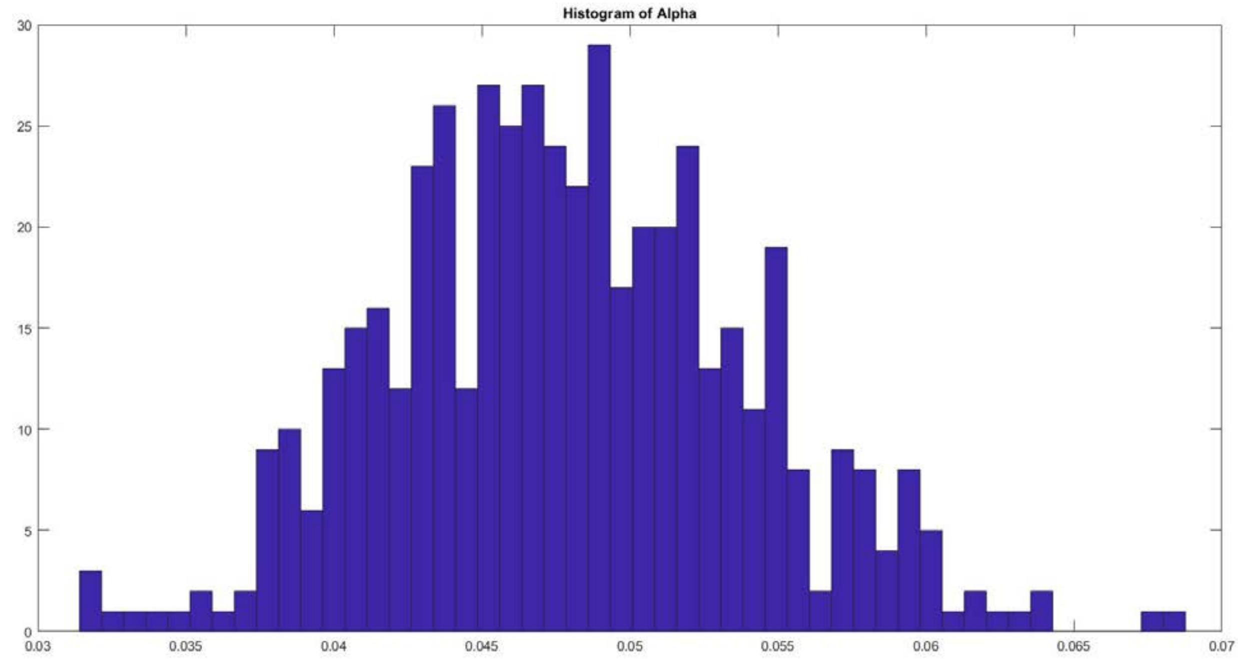


Figure 24: Model 5 - Histogram of α

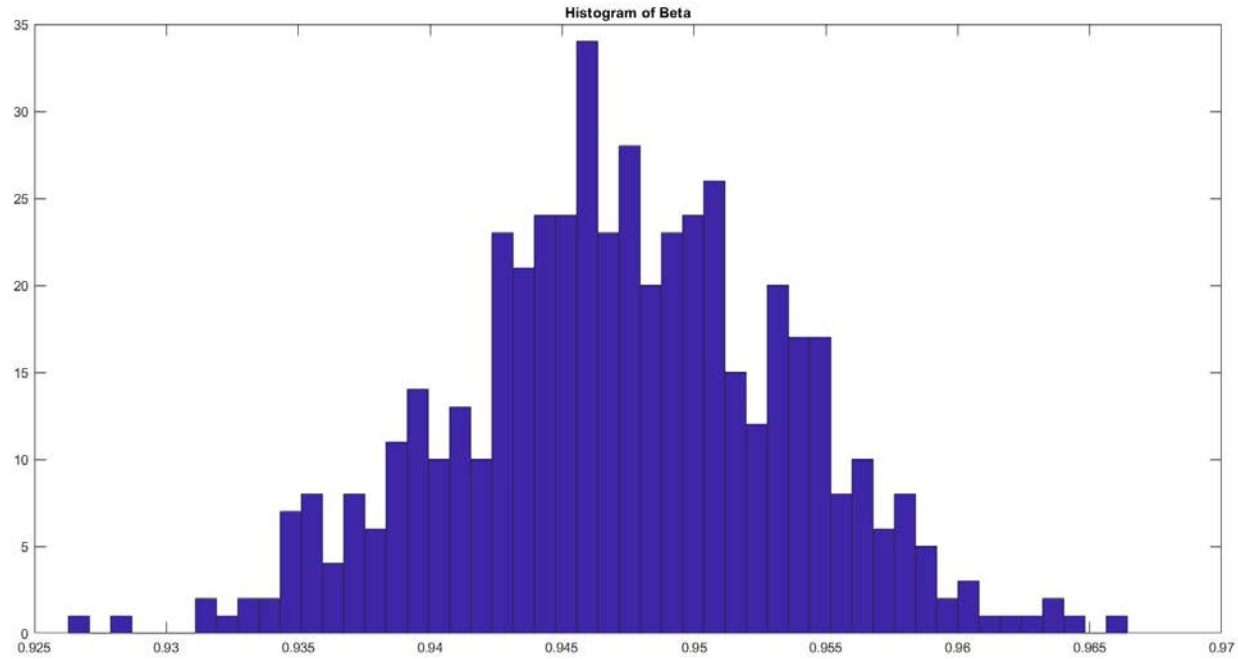


Figure 25: Model 5 - Histogram of β

The forecasted volatility for AstraZeneca, from the fifth model Aug-GARCH(1,1) with the lagged volume of news published is shown in Figure 26 and Figure 27.

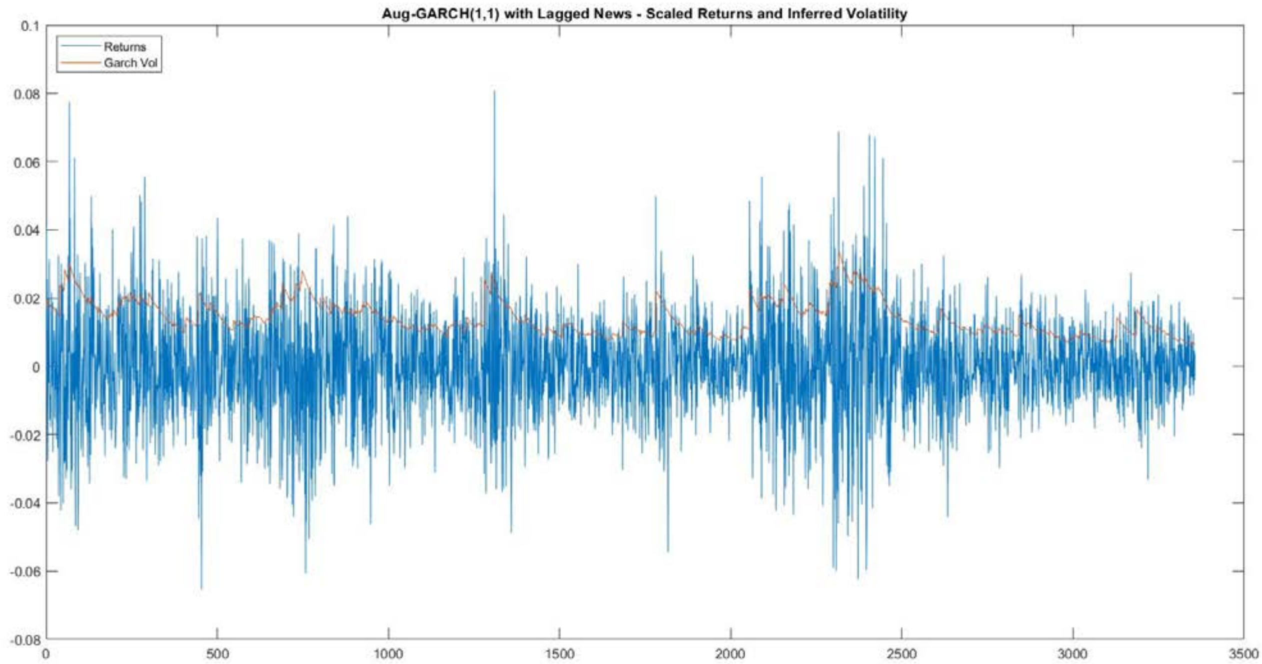


Figure 26: Inferred Volatility forecast and Scaled Returns of Model 5 for AstraZeneca

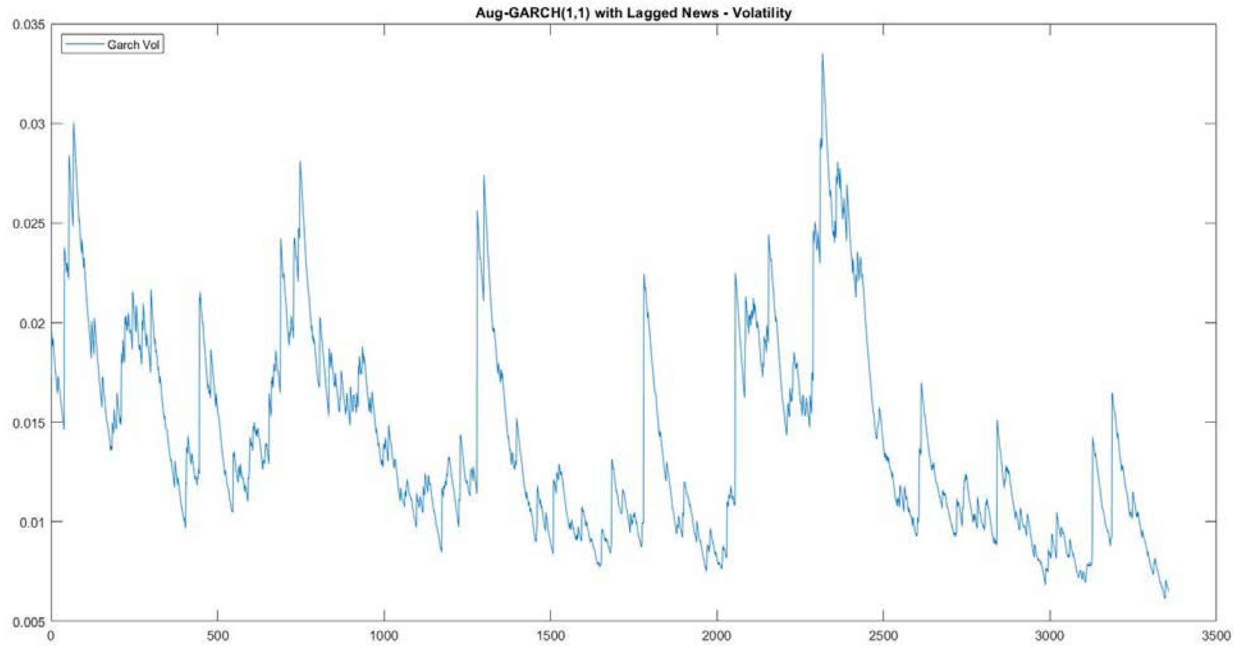


Figure 27: Volatility forecast of Model 5 for AstraZeneca returns

Figure 28 is the histogram of the Loglikelihood ratio between Model 1 and Model 5 for AstraZeneca after the Monte Carlo simulations have been run 1000 times.

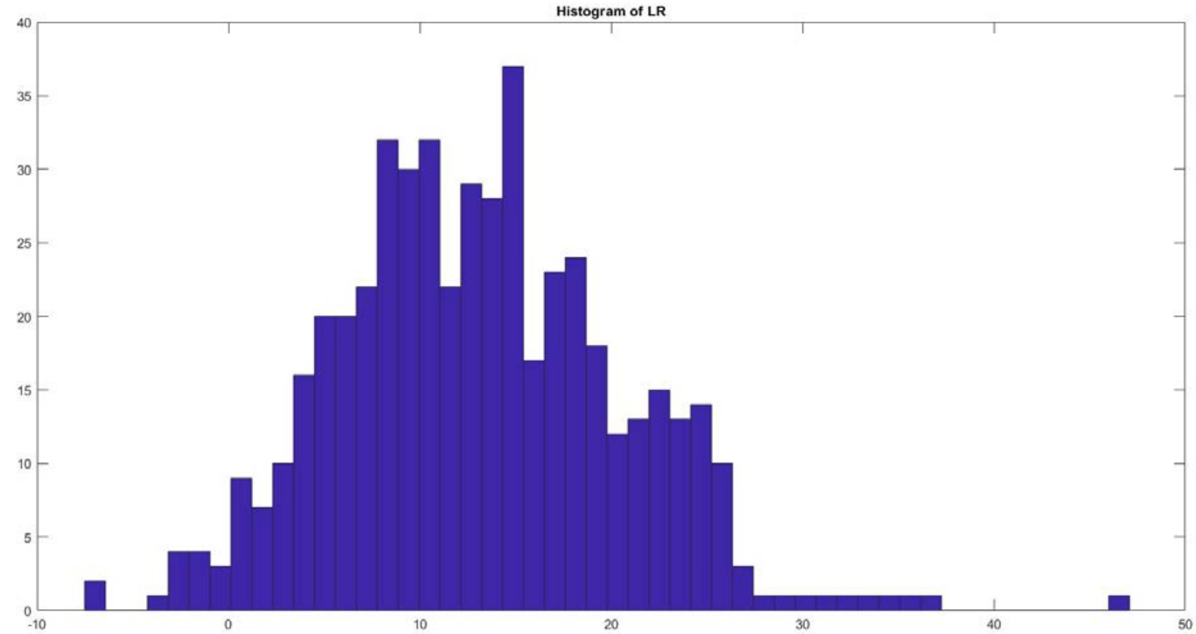


Figure 28: Histogram of Likelihood Ratio of Model 5

6.3 Conclusion

The results reported in Table 23 of Model 2, indicate that the inclusion of trading volume V_t, has compelling explanatory power in regards to the conditional volatility computed from the log returns. From the investigation of the chosen companies on the FTSE index, it is evident that when the contemporaneous trading volume is incorporated into the Aug-GARCH(1,1) model, there is a diminishing of GARCH effects for 3 of the 5 chosen companies, specifically AstraZeneca, Tesco and Vodafone. However, there is no indication of GARCH effects diminishing in Model 3, Aug-GARCH(1,1) with lagged daily trading volume, V_{t-1}. Additionally, the results of Model 4 and 5 are similar where it is shown that news volume, $news_t$ and Lagged News volume, $news_{t-1}$, do not remove GARCH effects either.

To compare the Aug-GARCH(1,1) models against the GARCH(1,1) model, Monte Carlo simulations were undertaken. This approach has been used in numerous studies such as Hansen and Lunde, (2005). In addition, the significance of the variance equations parameters was also examined. These parameters were estimated by QMLE.

Chapter 7: Summary and Future work

7.1 Part A: The impact of Press Releases Tone Upon Media Articles Sentiment

7.1.1 Summary

The first part of this research investigated if there are considerable discrepancies between the press releases of companies and the media articles which follow them, in terms of sentiment. This was achieved by examining the correlation between a corporation's disclosure sentiment and the corresponding media articles sentiment, through the textual analysis. The research results supported the assumption that financial journalists tend to publish a lowered tone for corporations' disclosures which sentiments are positive. This indicates that journalists are sceptical about positive information releases, in contrast to when the discourses are negative in tone. Overall, this part of the research contributes to the expanding body of studies, suggesting that the sentiment of information releases could be incremental information.

7.1.2 Future Work

It would be of interest to this study, that future research would seek to investigate how the financial press deciphers other information sources, such as social media services, which broadcasts financial content online. Additionally, future research could investigate different sources and if the information a source is releasing is consistently reliable. To elaborate, if a source is continuously publishing information in a positive tone about a certain corporation, whose stock values are steadily declining over time, does the information that is being published hold any value in weight?

7.2 Part B: The investigation of Aug-GARCH(1,1) Models with Volume of News Published

7.2.1 Summary

The second part of this research concentrates on the impact news has upon stock volatility. This was completed by implementing the data obtained from News Analytics, into Aug-GARCH(1,1) models. To compare the results of these models, Aug-GARCH(1,1) models were undertaken with trading volume, in addition to a GARCH(1,1) model. The dataset used in this study consisted of 5 chosen companies over the time period January 1st, 2000 until December 31st, 2012. The news dataset of the 5 chosen companies were extracted from *Factiva*, whilst the stock dataset was collected from *Bloomberg*. There is a limited body of work which focuses on the investigating the volume of news releases influence on stock volatility. As mentioned previously, the research of Janssen (2004) focuses on the influence the frequency of information releases has on index volatility. This study differed from Janssen's (2004) as the goal was to examine the effect on stock volatility. From the results obtained on the empirical study of FTSE companies, it was demonstrated that Model 2 removed GARCH effects for 3 of the 5 selected companies. However, Models 4 and 5, Aug-GARCH(1,1) models with news volume, $news_t$, experienced difficulties removing these effects.

Monte Carlo simulations with the log likelihood ratio, was used in this study to test the models. Through this test, the following conclusions were reached:

- For Model 2, 4 out of the 5 chosen corporations reject the null hypothesis with a confidence level of 1%.
- For Model 3, all of the 5 chosen corporations accept the null hypothesis with a confidence level of 1%.
- For Model 4, the results recorded show that for all the 5 chosen corporations, the null hypothesis is rejected with a confidence level of 1%.
- For Model 5, the results recorded show that for all the 5 chosen corporations, the null hypothesis is rejected with a confidence level of 1%.

Using QMLE, the times series models were estimated, using the daily prices models of the chosen FTSE corporations. The models were calibrated using QMLE Methods, through the software *MATLAB*.

7.2.2 Future Work

This research demonstrates that examining the influence of news publishing volume upon stock volatility, is much more complex than it initially appeared. The results obtained do not provide strong indications that the number of articles published has a distinct impact on volatility. There are various reasons why this may have occurred, which are as follows:

- Further news analysis is necessary upon the dataset;
- The discrepancies in volatility from the '2008 Financial Crisis' were incorporated into the stock dataset, whilst the volume of news published did not vary as vastly in comparison during the whole period.

This study could be treated as a preliminary investigation in the area of news impact evaluation upon stock volatility. In addition, could potentially form a foundation to suggest new directions for future studies. Future work could focus on adapting the Aug-GARCH(1,1) to include jumps. This would act as a continuation on from the research of Jorion, (1988), who examined a mixed jump-diffusion process, thus Models 4 and 5 could be adapted to investigate the impact news volume could have upon the intensity of jumps occurring in volatility.

Appendix A – Graphs – Sentiment Data

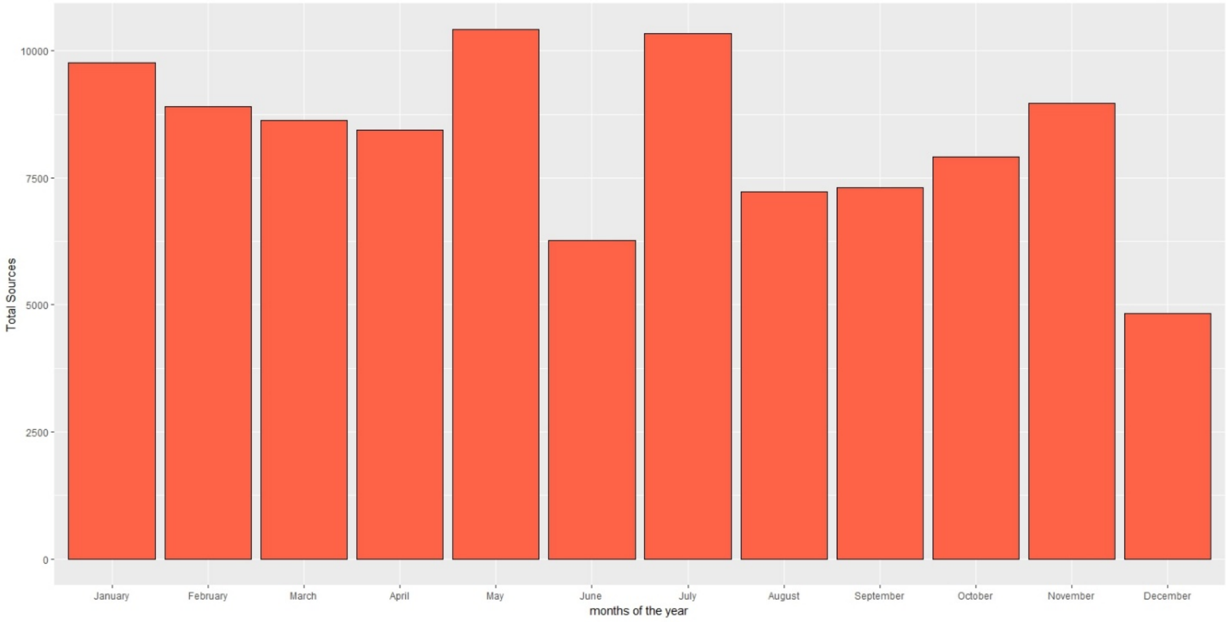


Figure 29: Bar Graph illustrating the no. of Total Sources published by Months of the Year

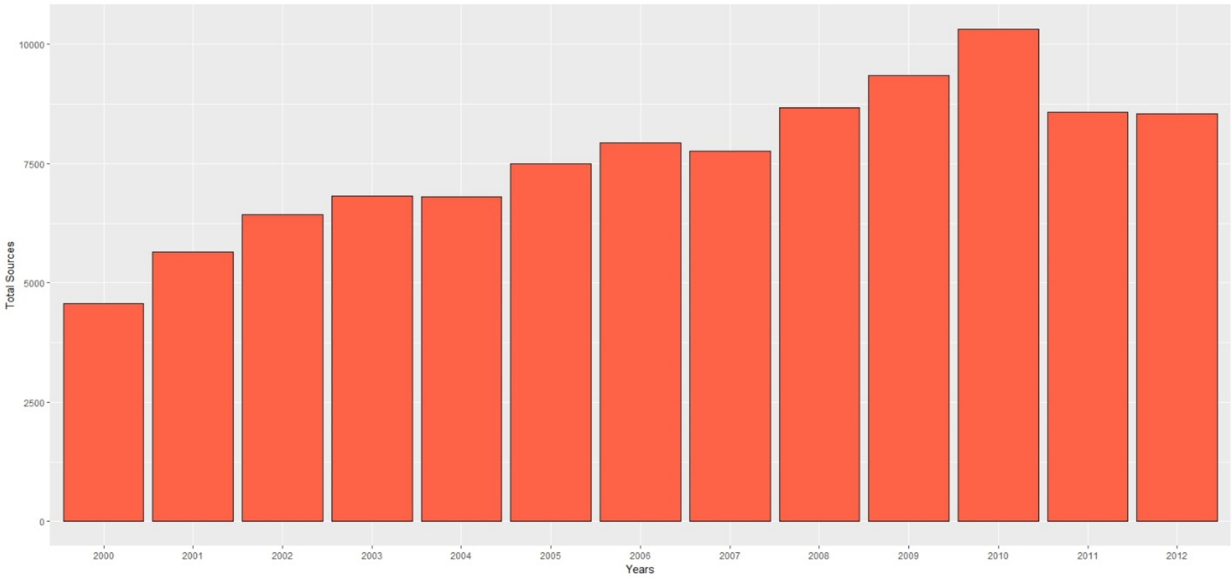


Figure 30: Bar Graph illustrating the no. of total sources published by Year

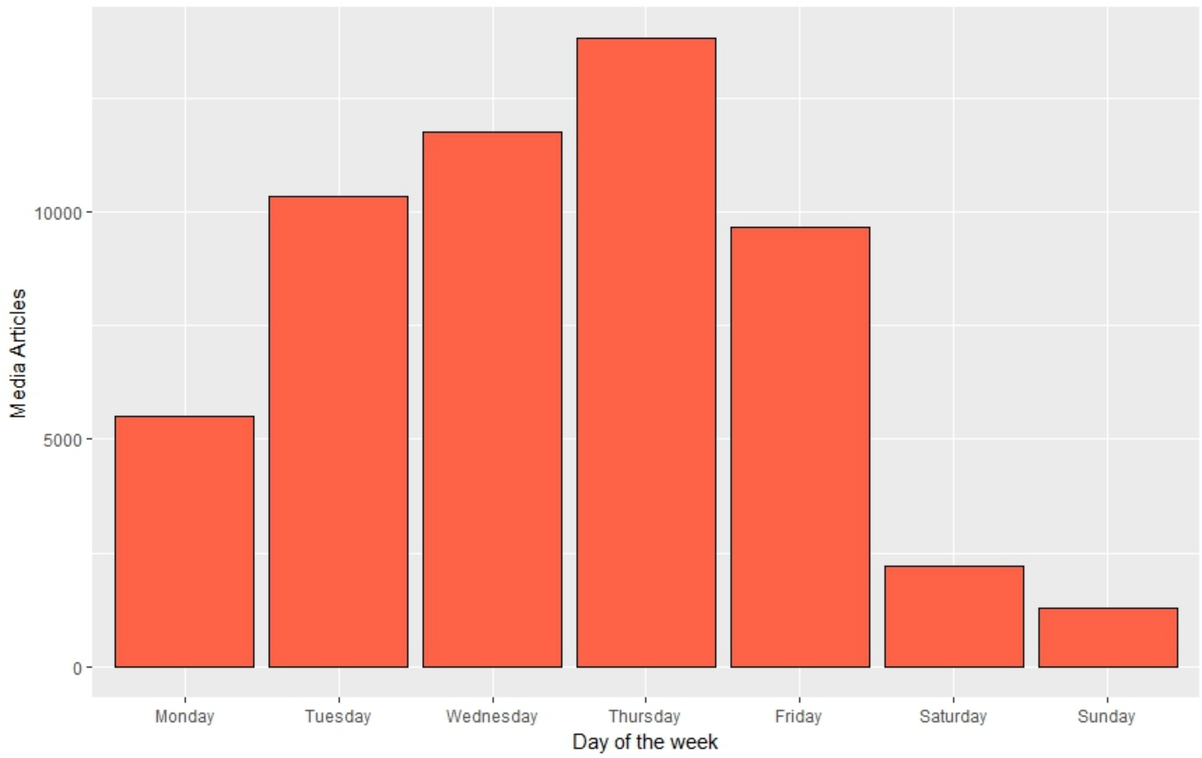


Figure 31: Bar Graph illustrating the no. of total sources published by Days of the Week

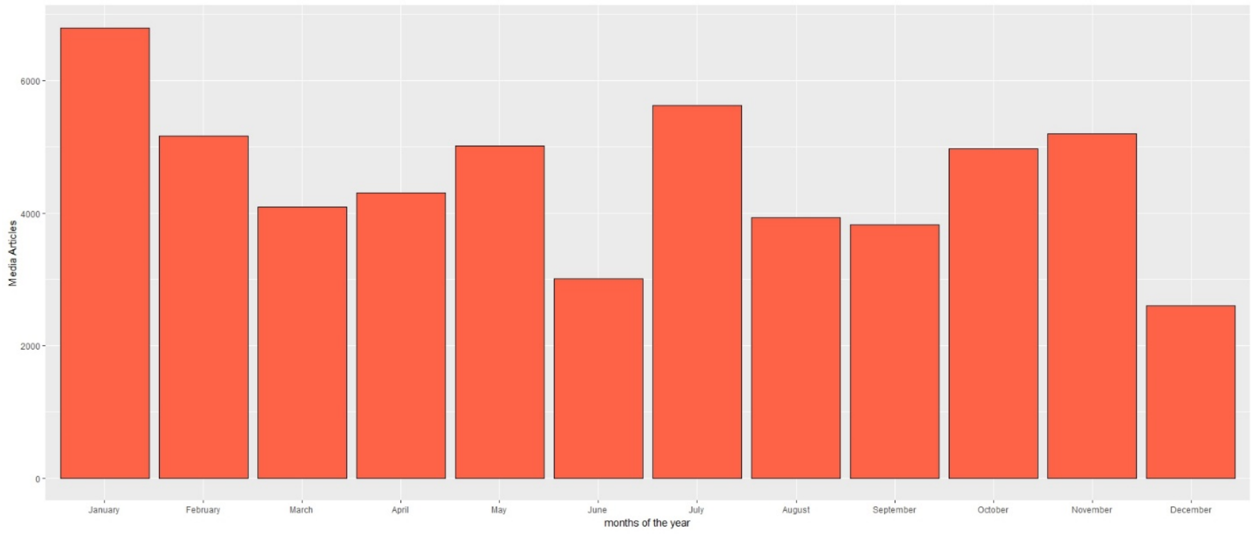


Figure 32: Bar Graph illustrating the no. of Media Articles published by Months of the Year

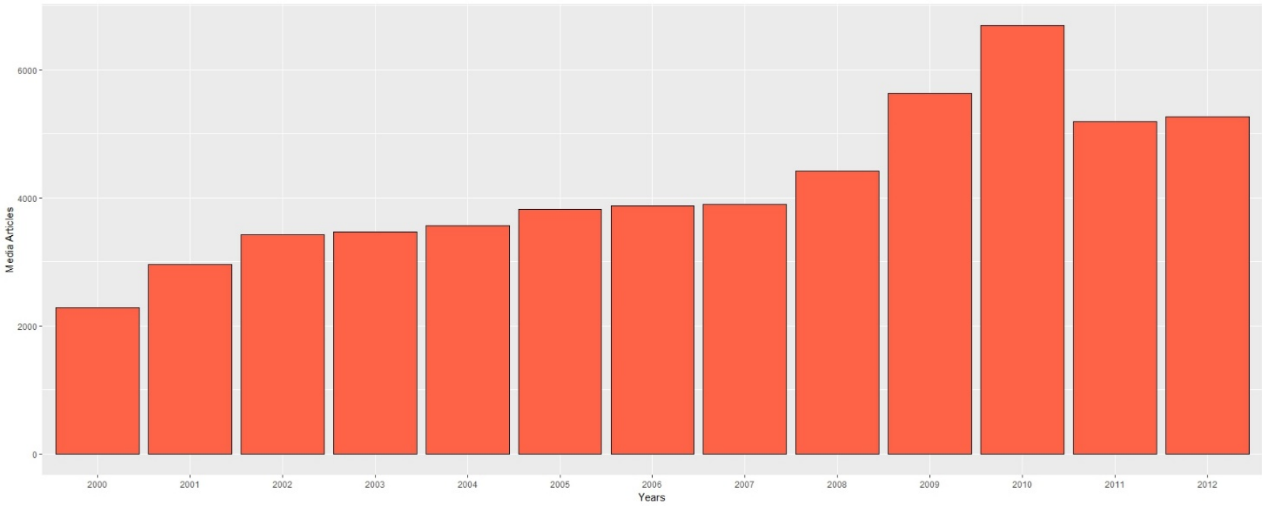


Figure 33: Bar Graph illustrating the no. of Media Articles published by Year

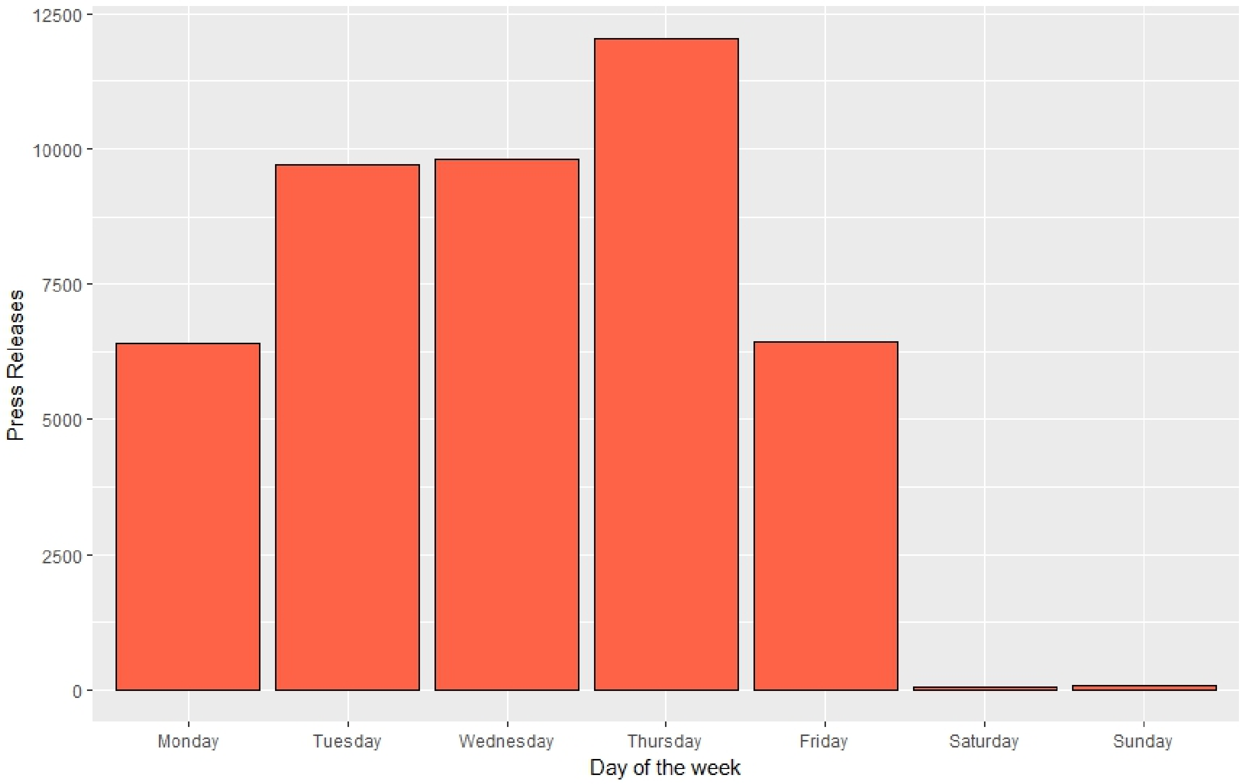


Figure 34: Bar Graph illustrating the no. of Media Articles distributed by Days of the Week

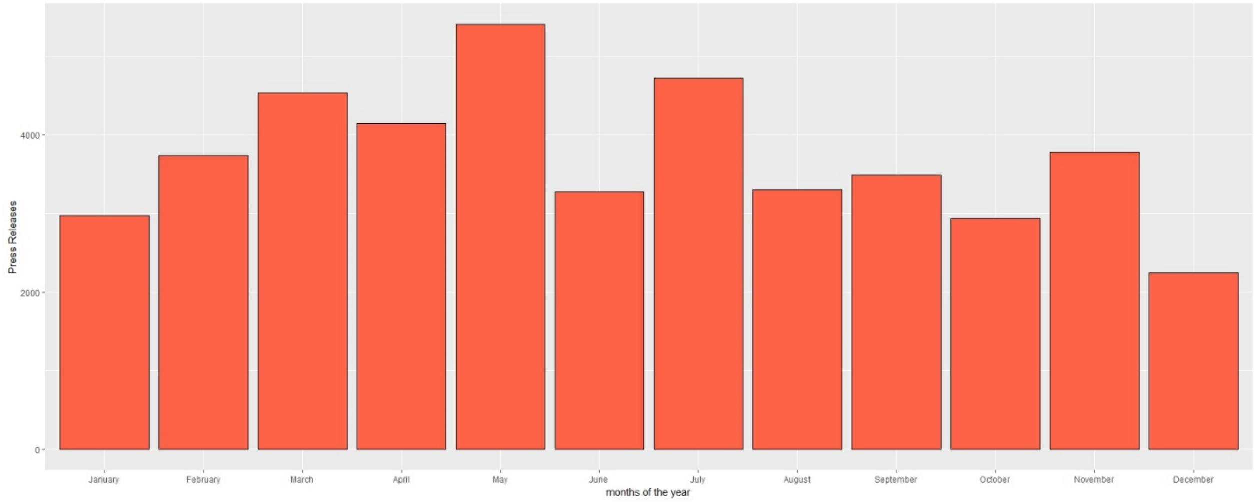


Figure 35: Bar Graph illustrating the no. of Press Releases published by Months of the Year

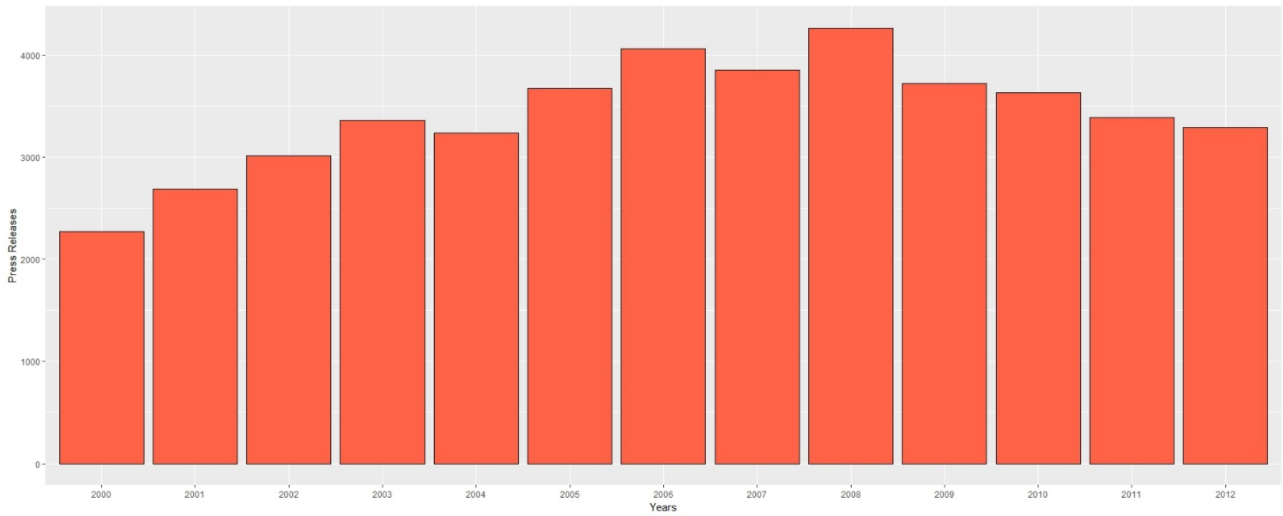


Figure 36: Bar Graph illustrating the no. of Press Releases published by Year

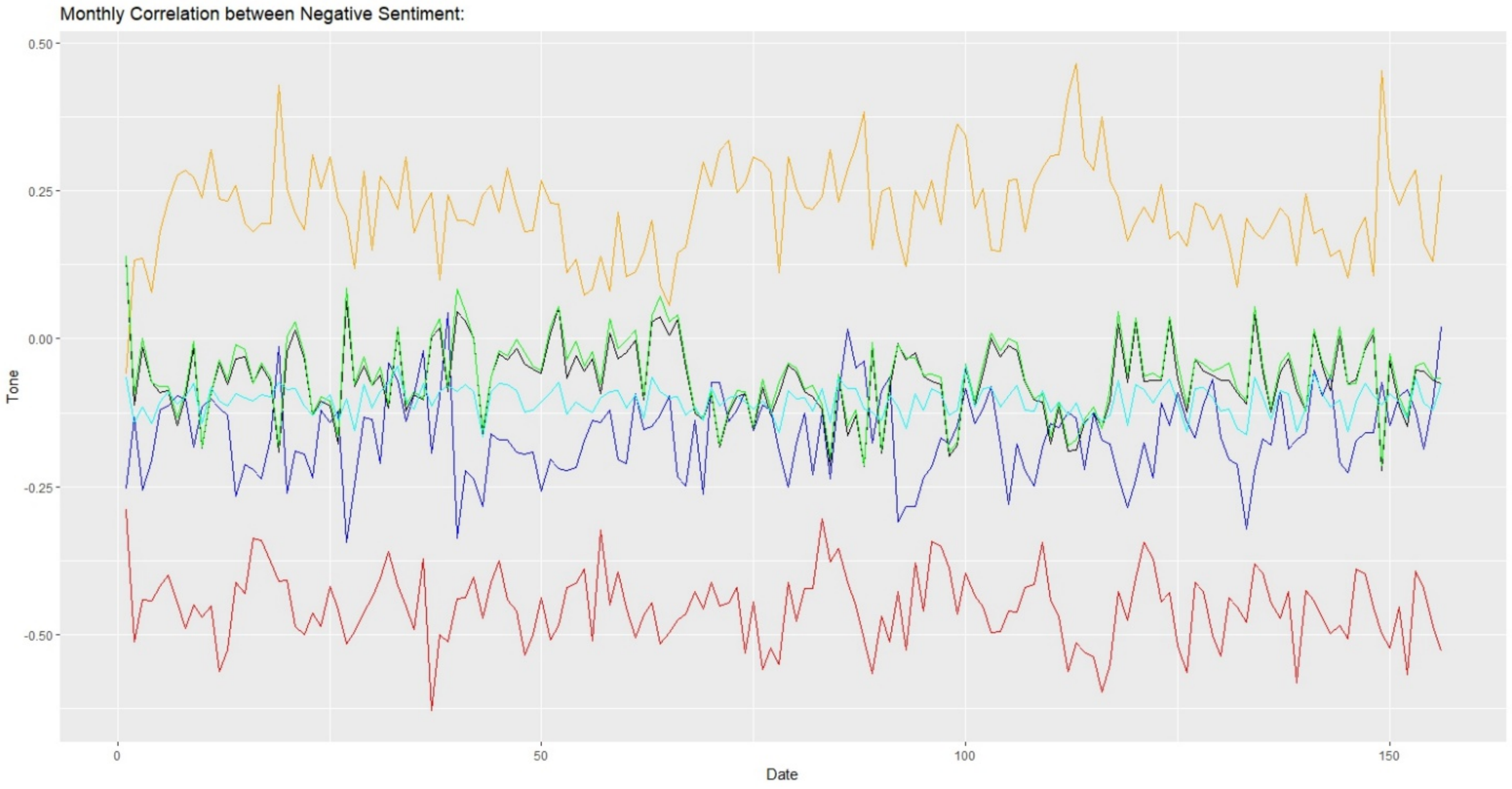


Figure 37: Monthly Correlation between Negative Sentiment

Note: The Legend for the above Figure can be found in Table 6. (page 36)

Appendix B – Graphs – Stock Data

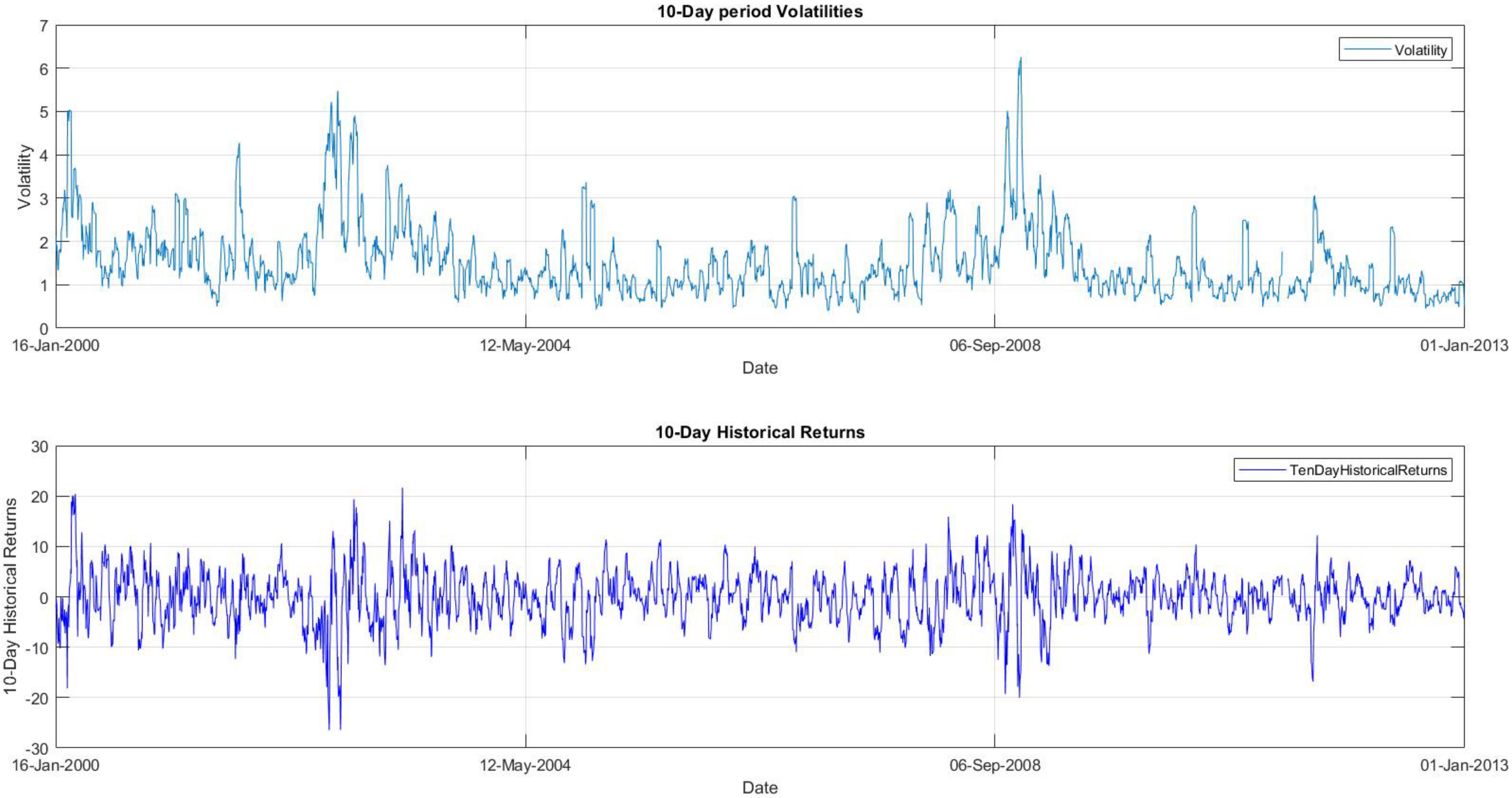


Figure 38: Comparison between AstraZeneca's 10-Day Period Volatilities and 10 Day Historical Returns

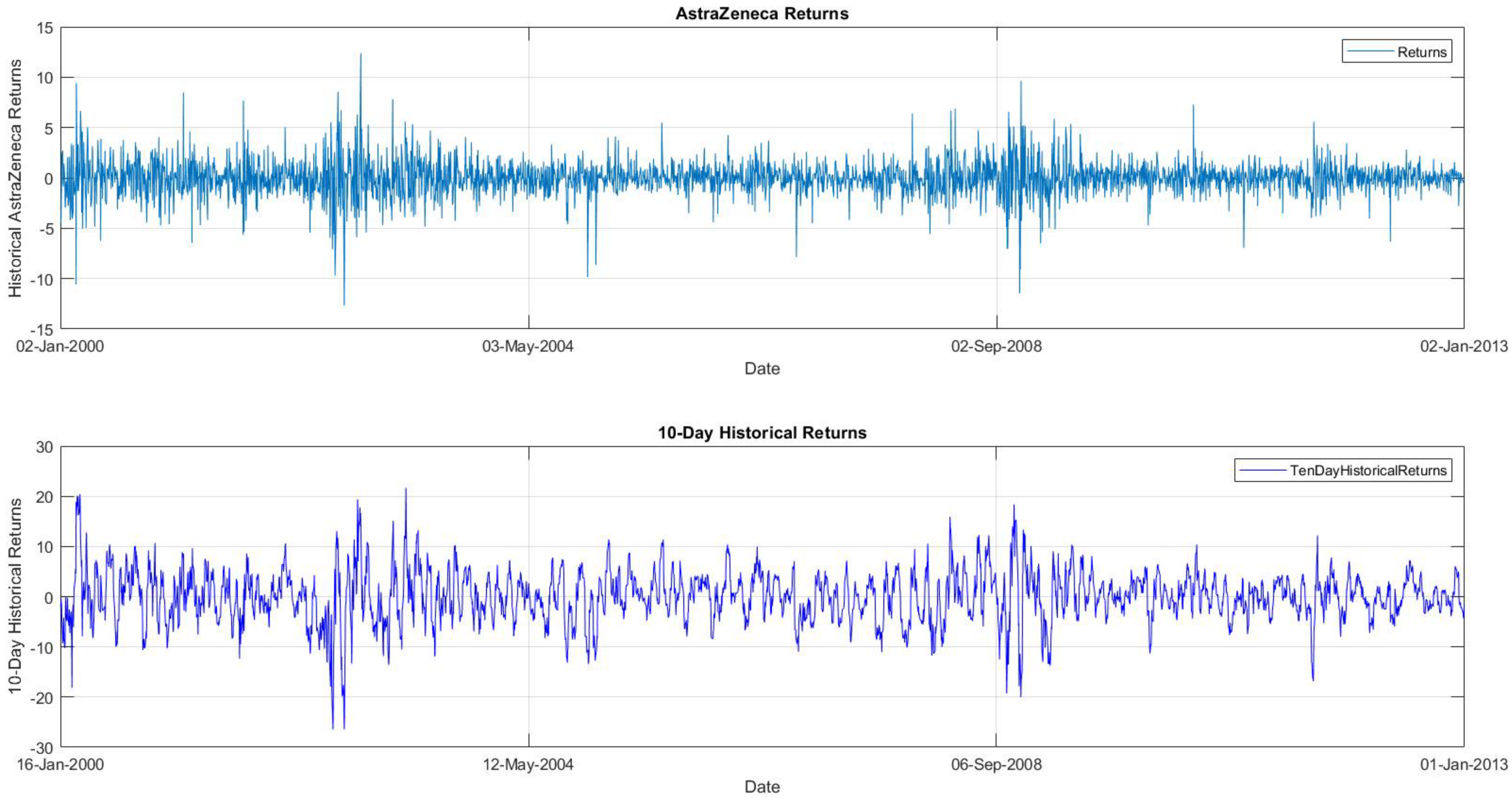


Figure 39: Comparison between AstraZeneca's Daily Historical Returns and 10-Day Historical Returns

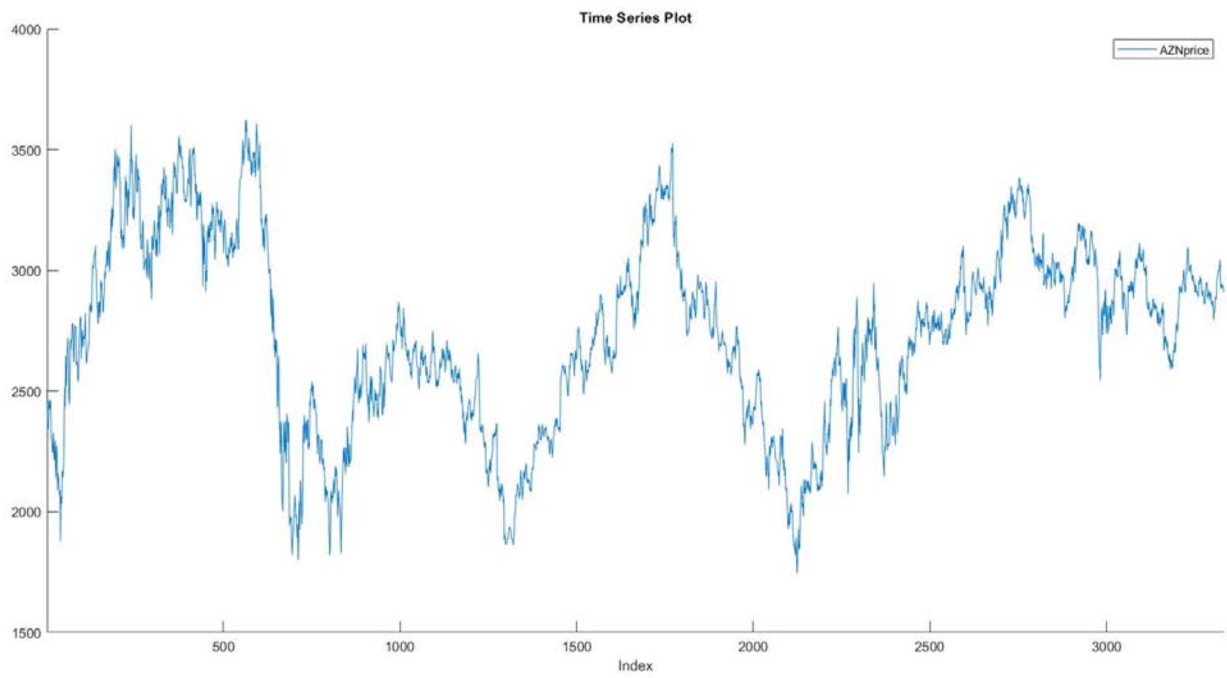


Figure 40: Historical Movement of AstraZeneca's Daily Closing Prices

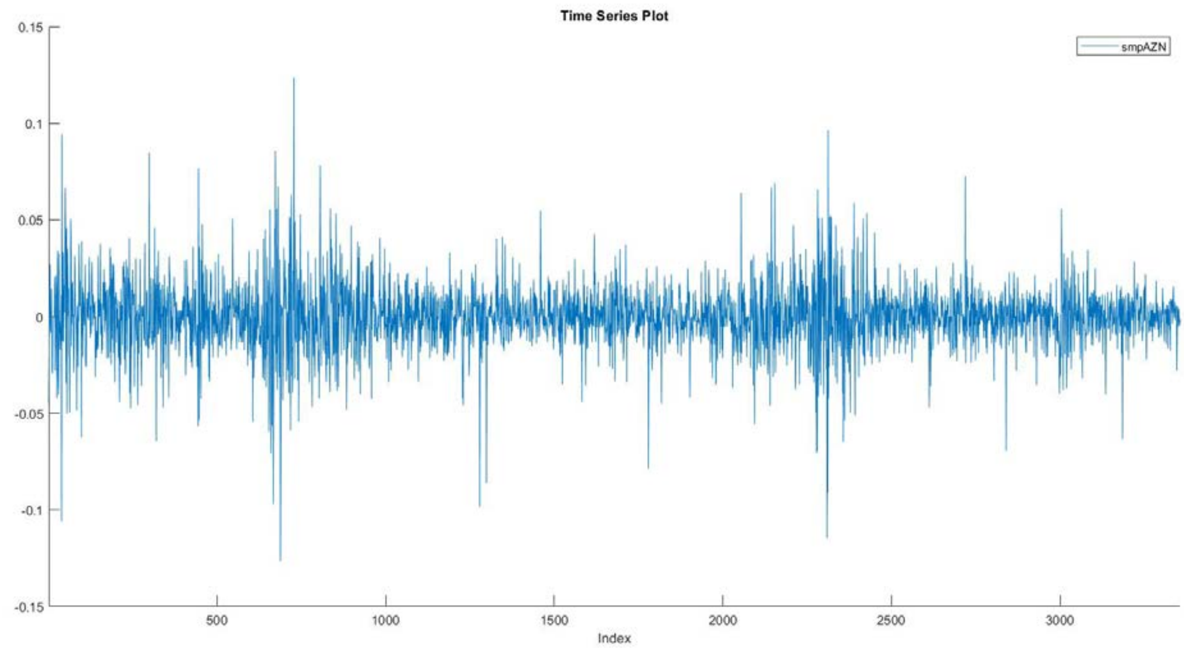


Figure 41: Historical Movement of AstraZeneca's Daily Returns

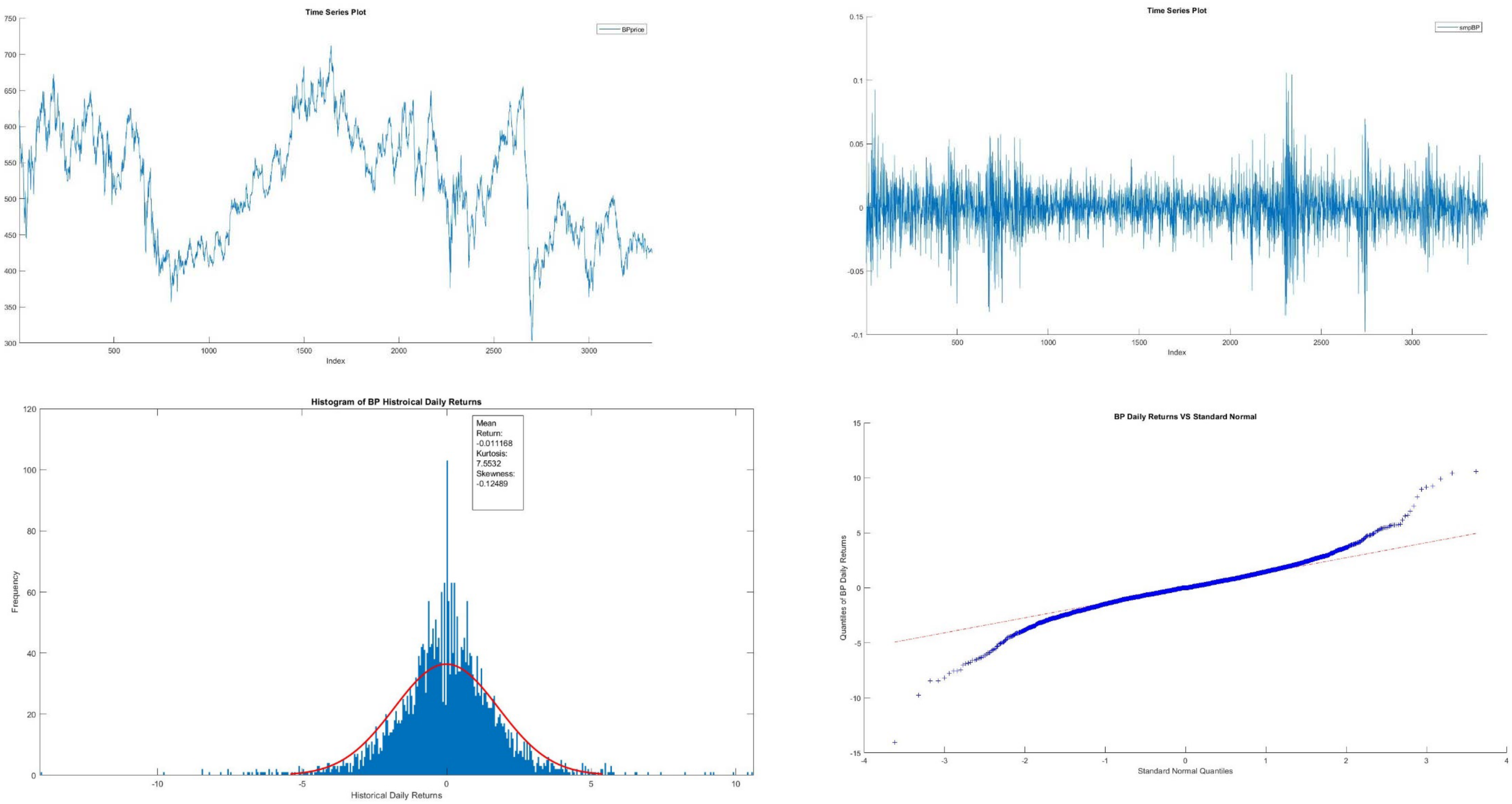


Figure 42: Historical Movement of BP's Daily Closing Prices, Daily Returns, Histogram of Daily Returns and QQ Plot of Daily Returns and Standard Normal Quantiles

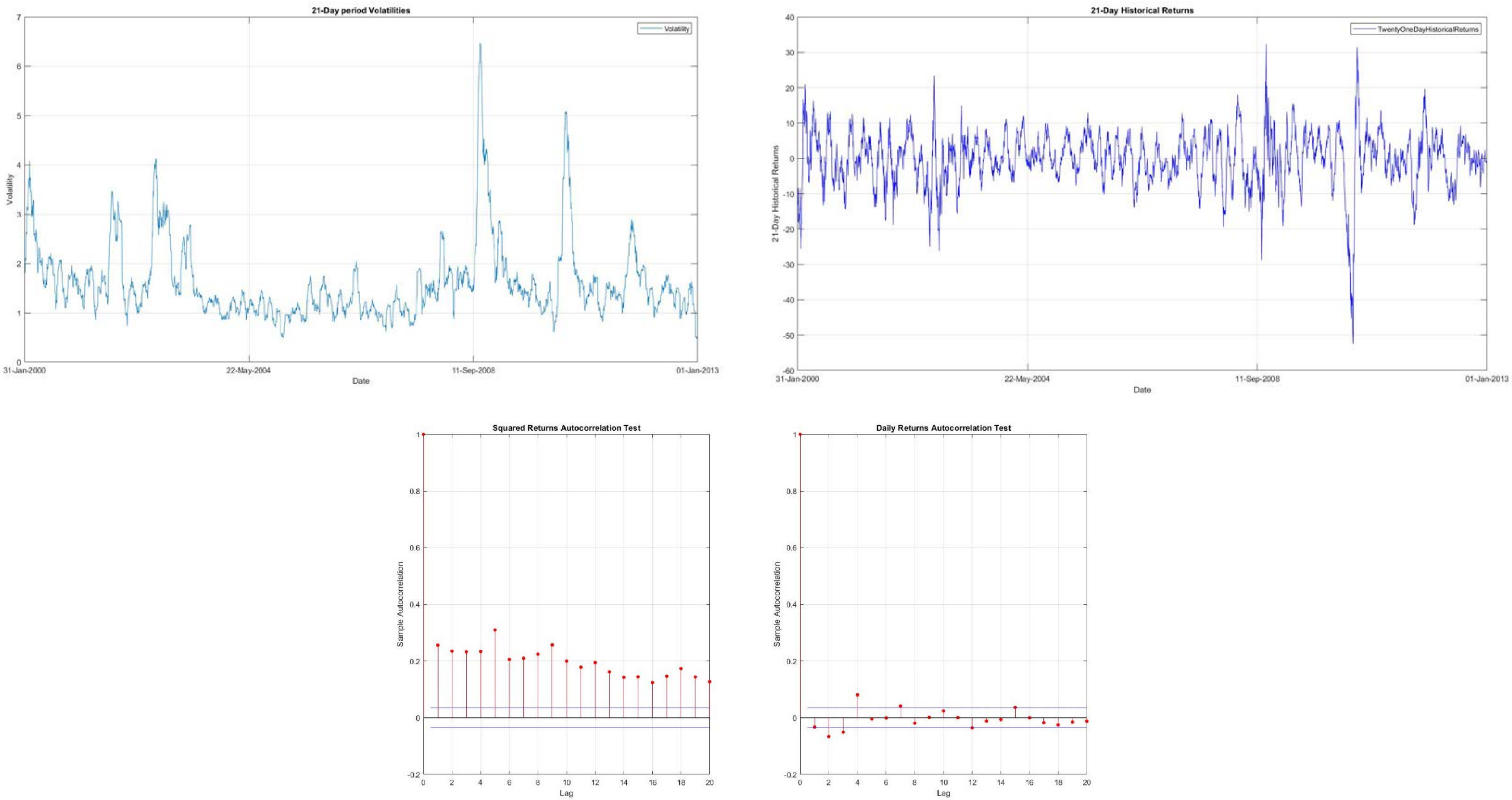


Figure 43: Historical Movement of BP's 21-day Volatility, 21-day Returns and Autocorrelation for Squared Returns and Returns

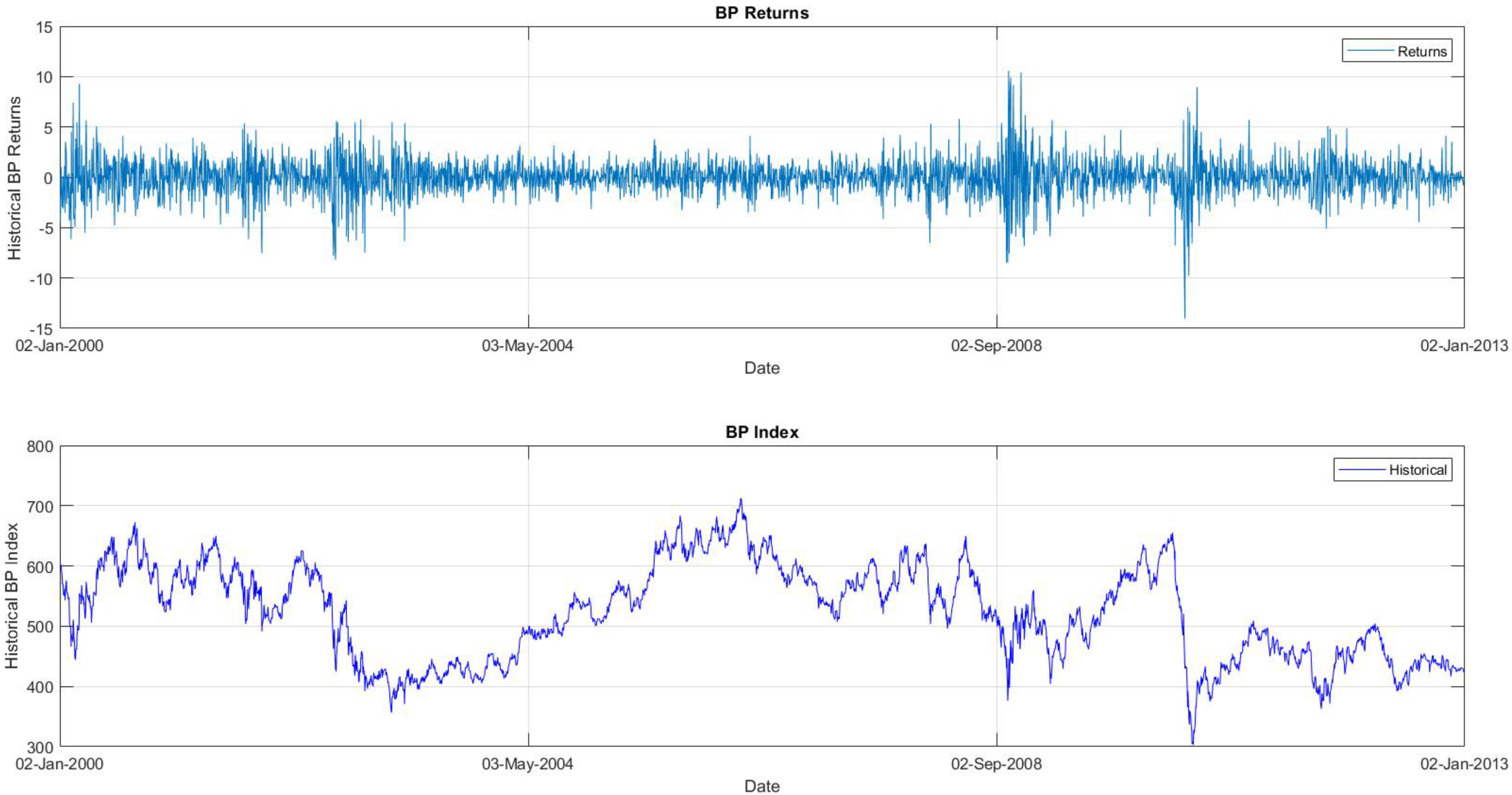


Figure 44: Comparison between BP's Closing Daily Prices and Historical Returns

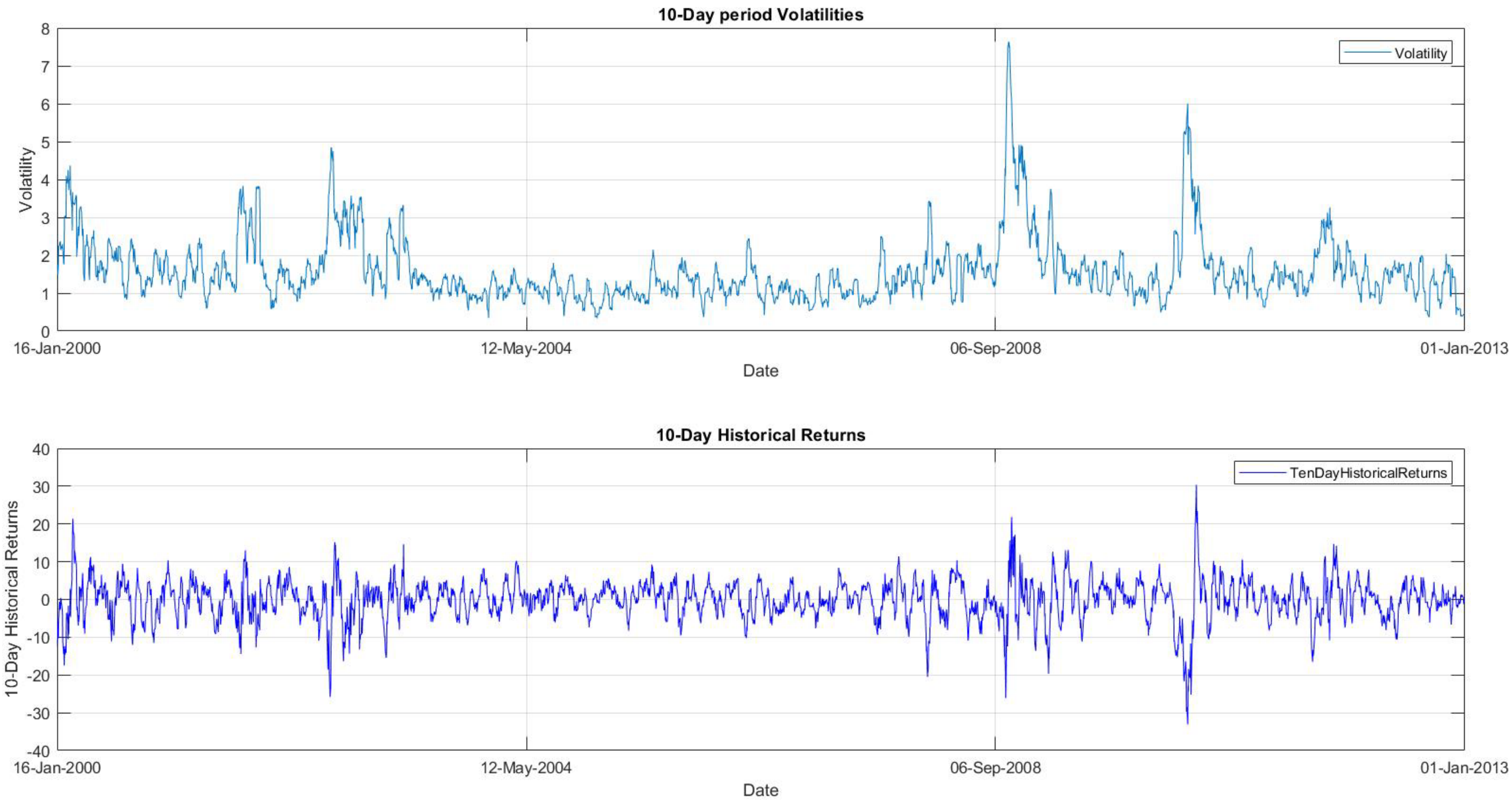


Figure 45: Comparison between BP's 10-Day Period Volatilities and 10 Day Historical Returns

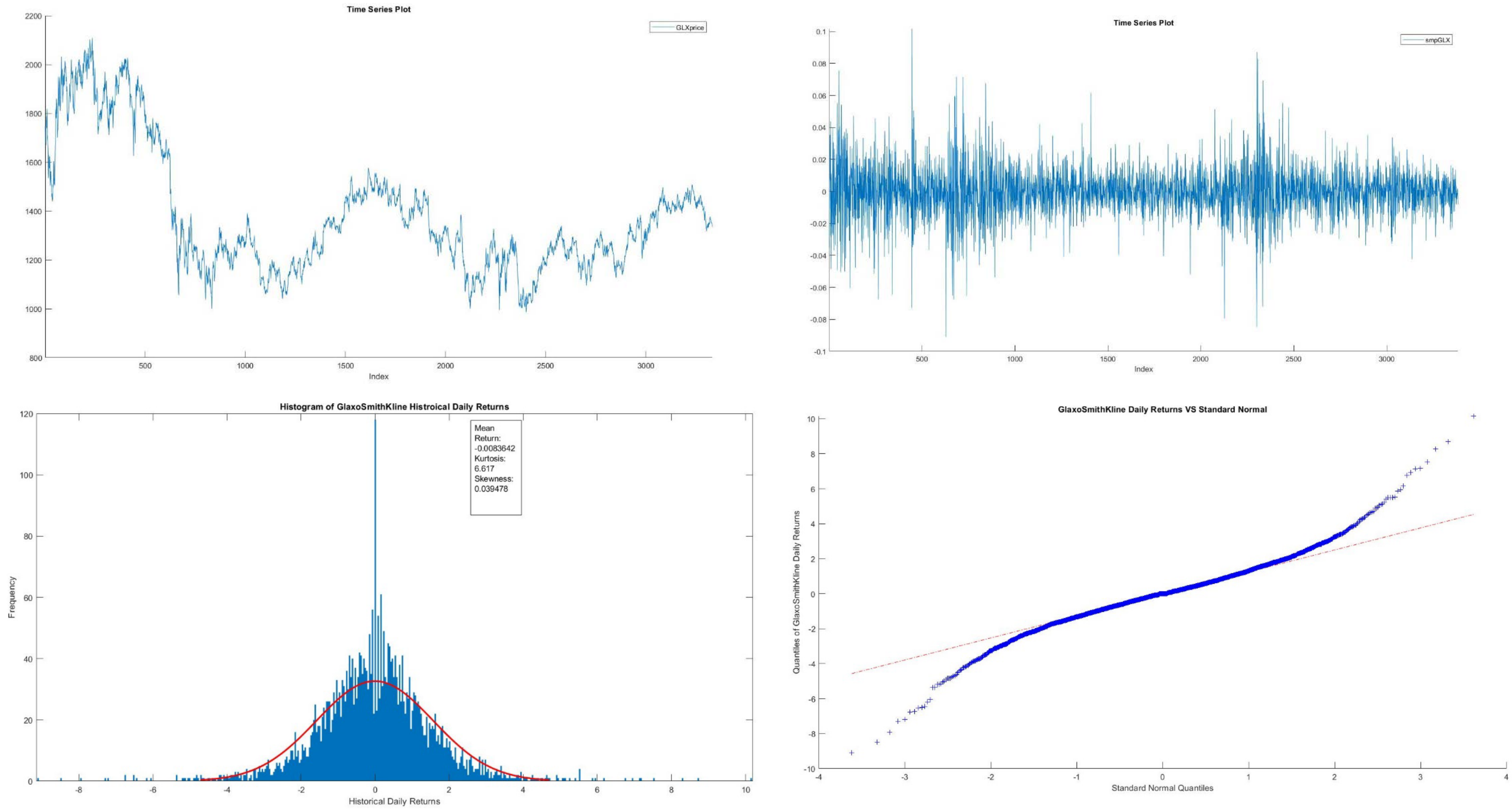


Figure 46: Historical Movement of GlaxoSmithKline's Daily Closing Prices, Daily Returns, Histogram of Daily Returns and QQ Plot of Daily Returns and Standard Normal Quantiles

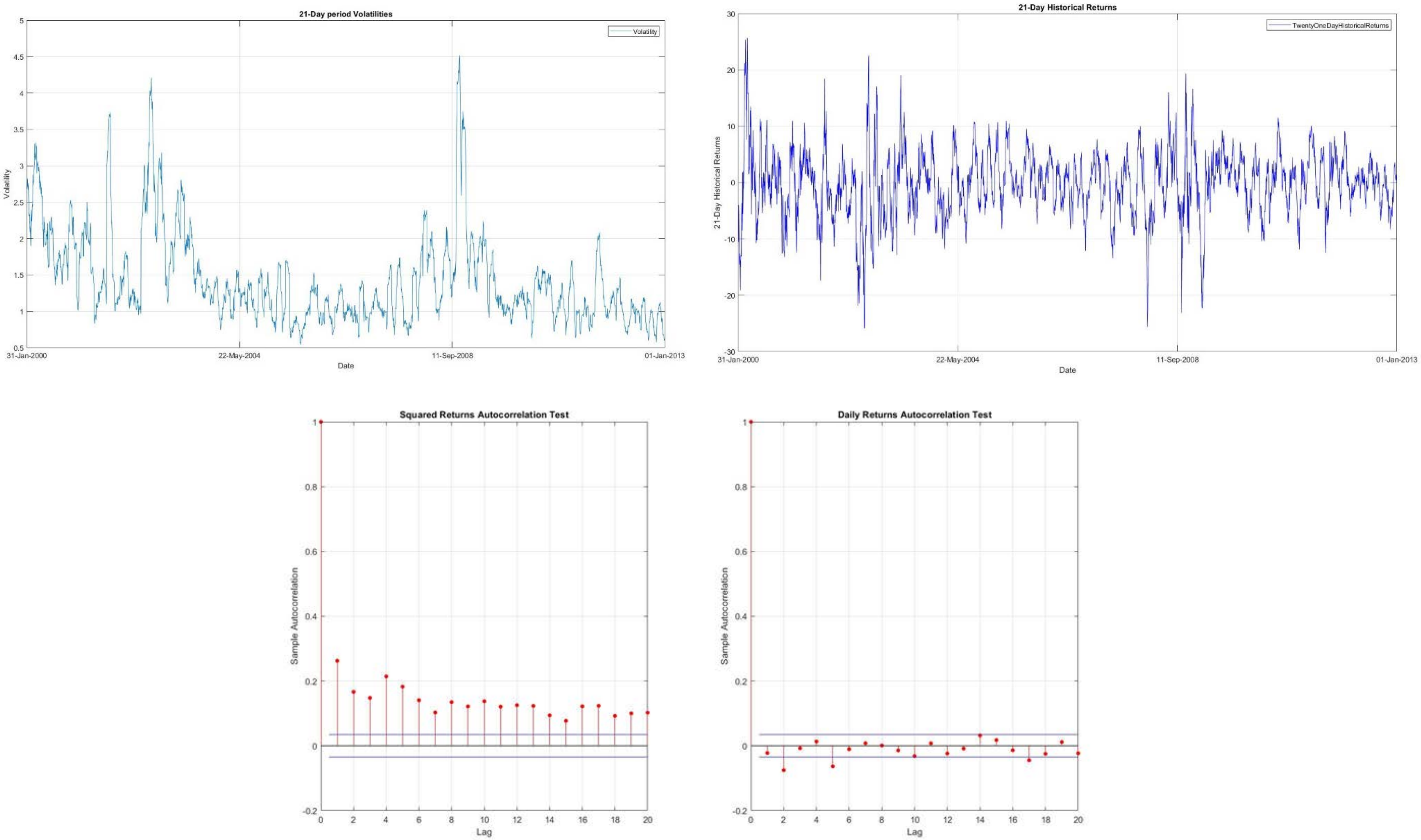


Figure 47: Historical Movement of GlaxoSmithKline's 21-day Volatility, 21-day Returns and Autocorrelation for Squared Returns and Returns

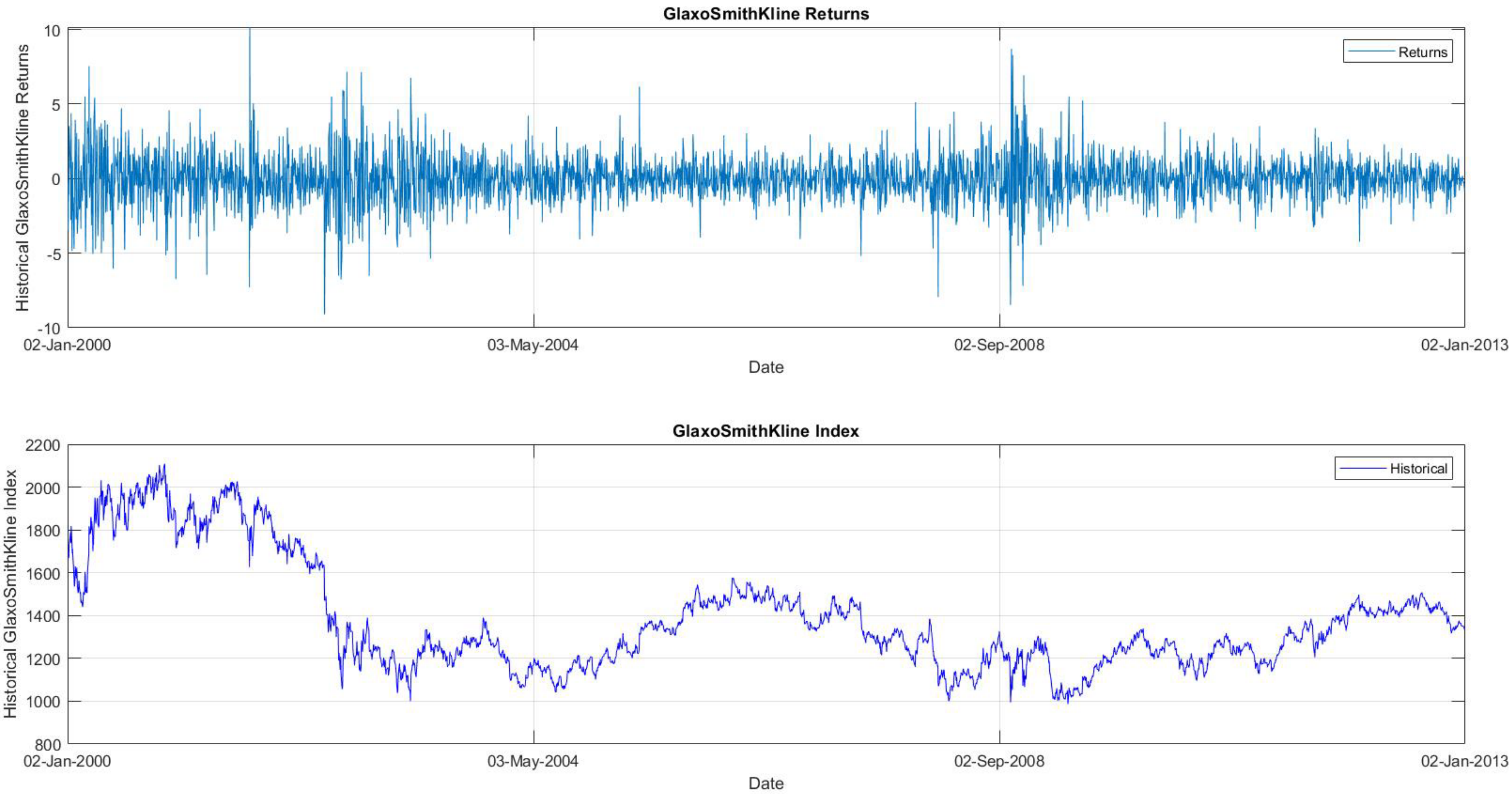
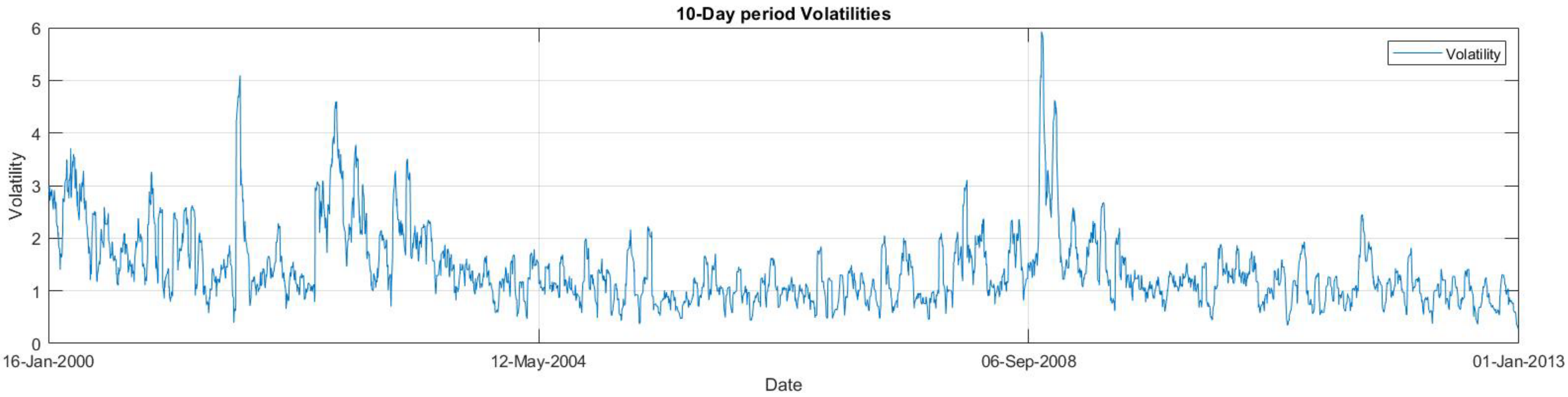


Figure 48: Comparison between GlaxoSmithKline's Closing Daily Prices and Historical Returns



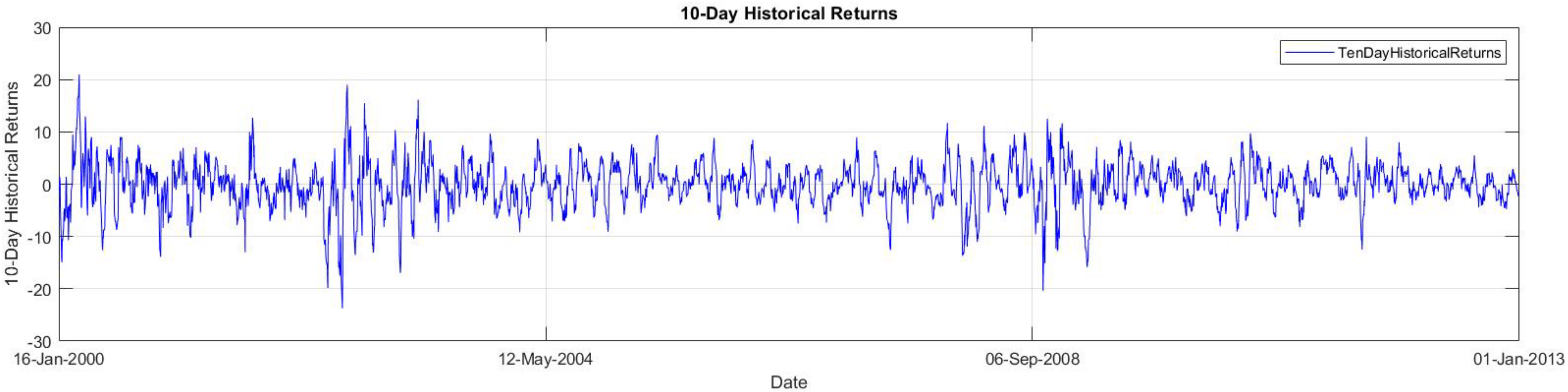


Figure 49: Comparison between GlaxoSmithKline's 10-Day Period Volatilities and 10 Day Historical Returns

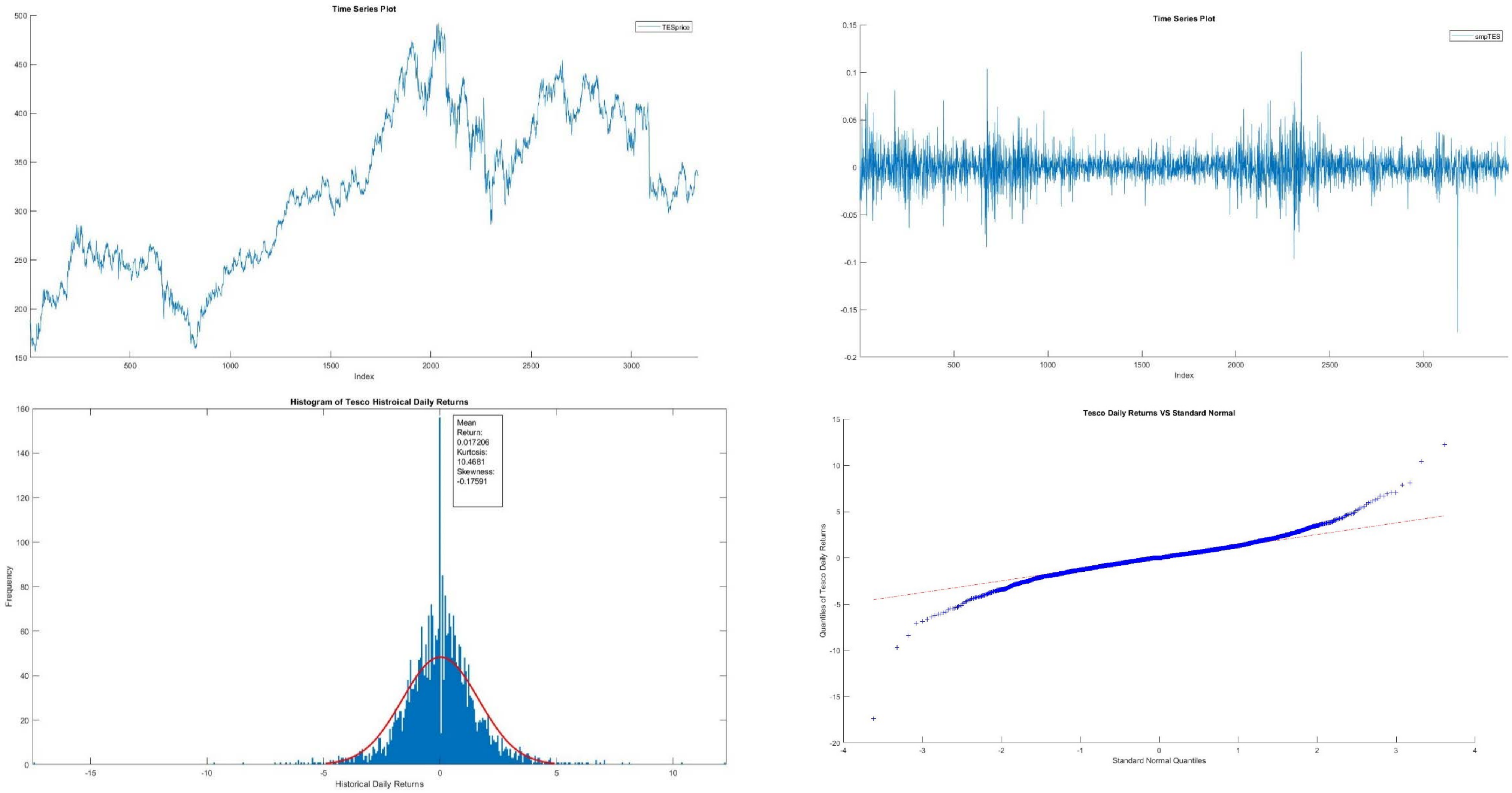


Figure 50: Historical Movement of Tesco's Daily Closing Prices, Daily Returns, Histogram of Daily Returns and QQ Plot of Daily Returns and Standard Normal Quantiles

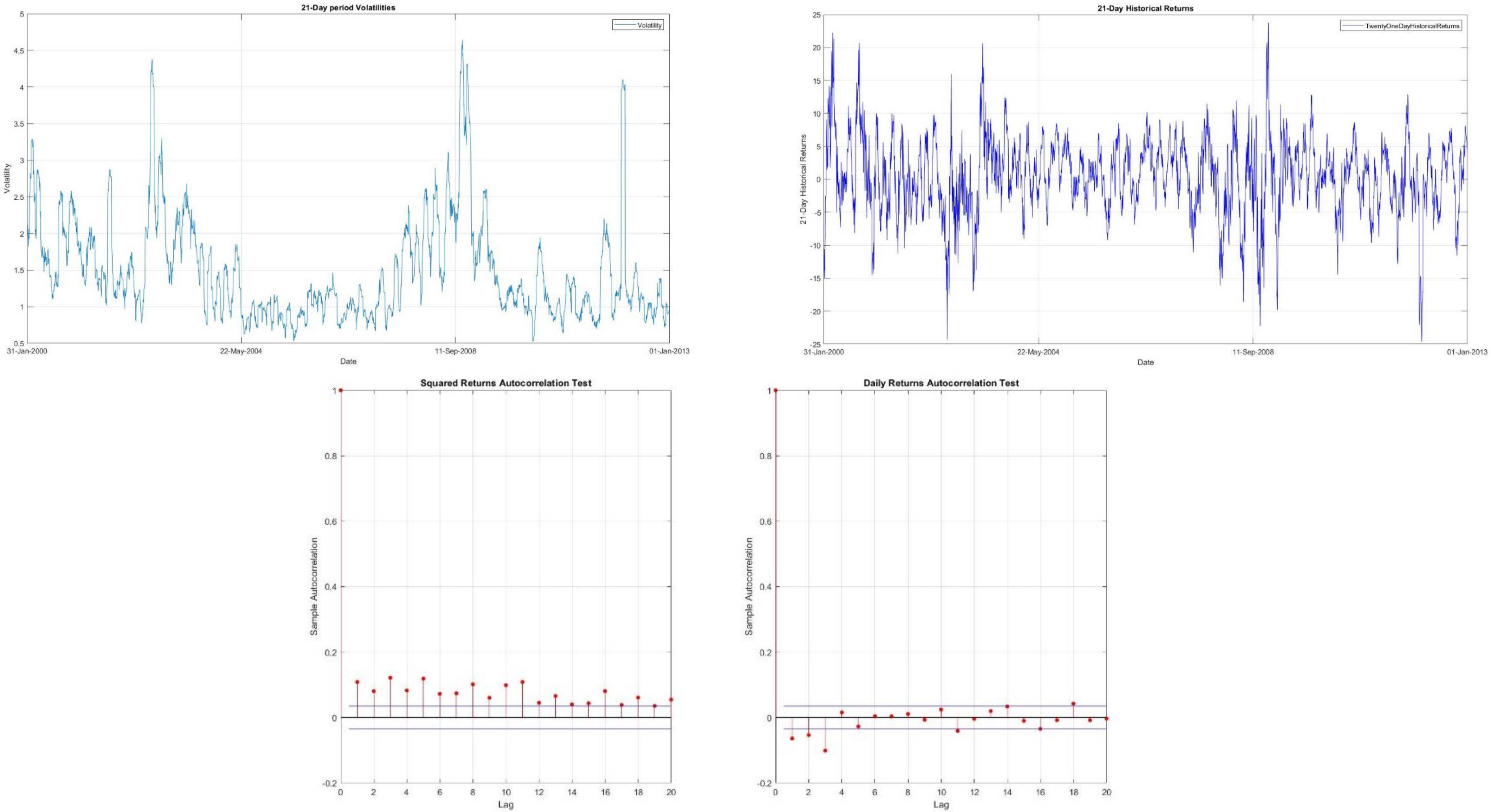


Figure 51: Historical Movement of Tesco's 21-day Volatility, 21-day Returns and Autocorrelation for Squared Returns and Returns

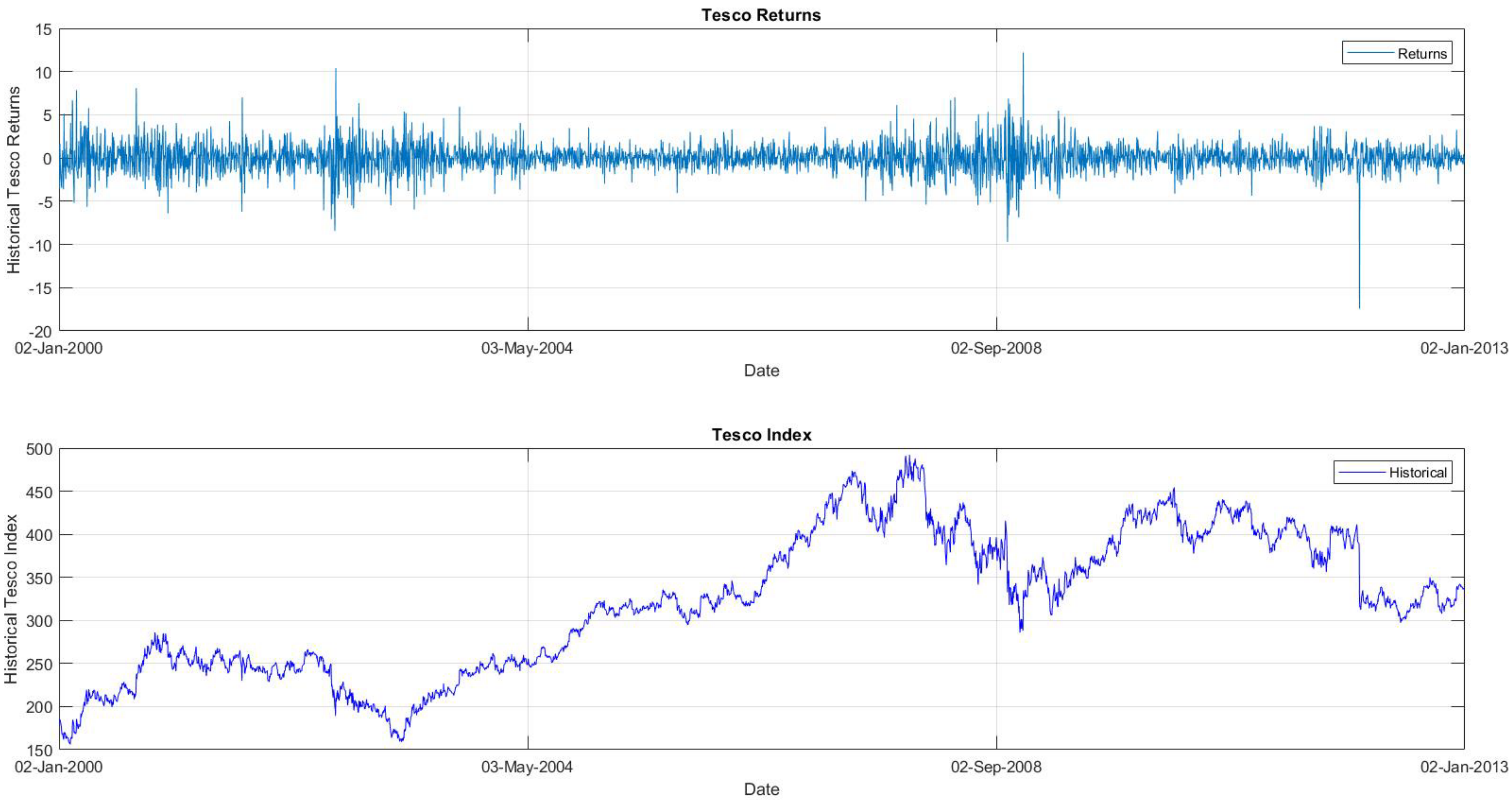


Figure 52: Comparison between Tesco's Closing Daily Prices and Historical Returns

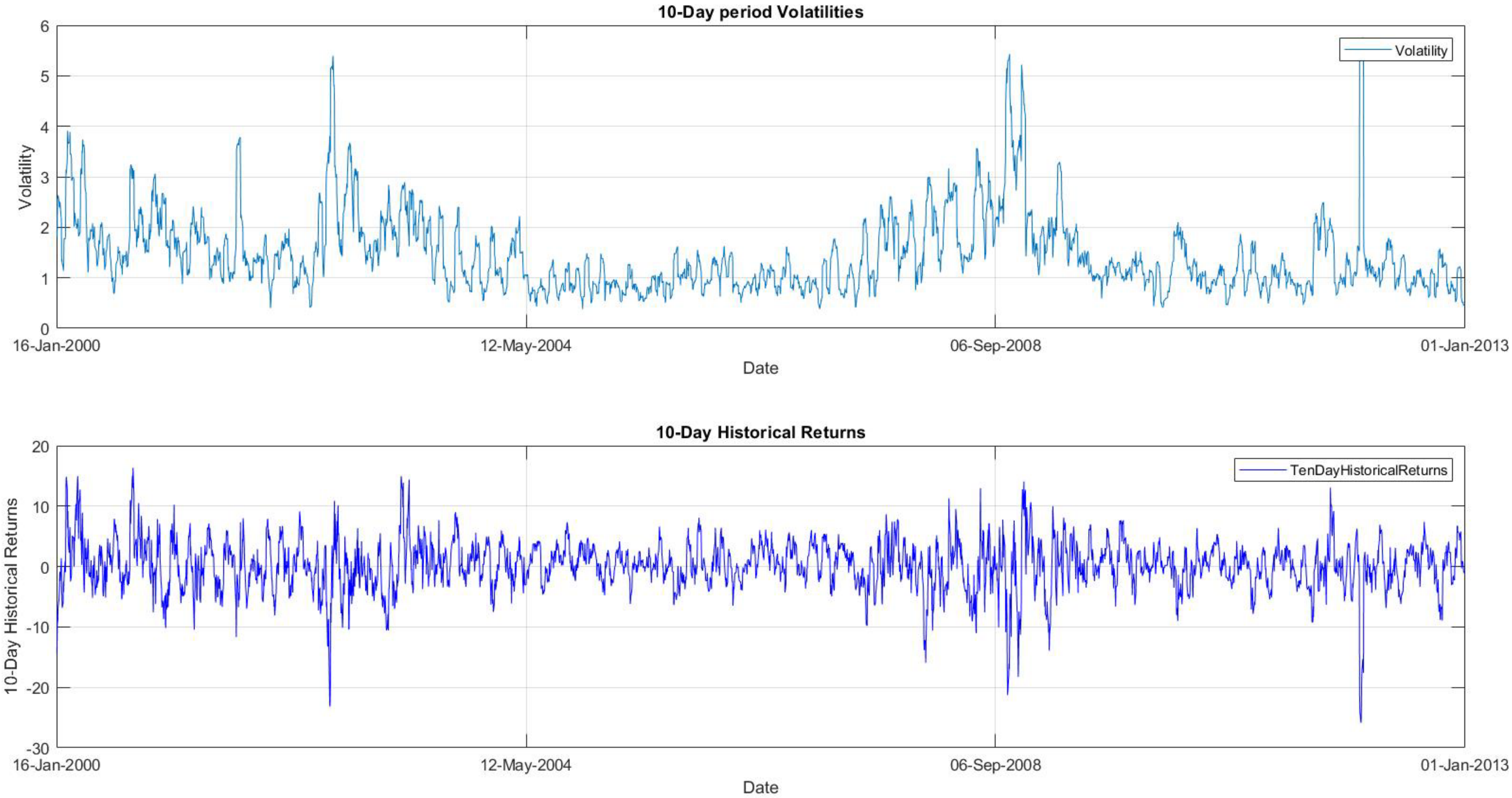


Figure 53: Comparison between Tesco's 10-Day Period Volatilities and 10 Day Historical Returns

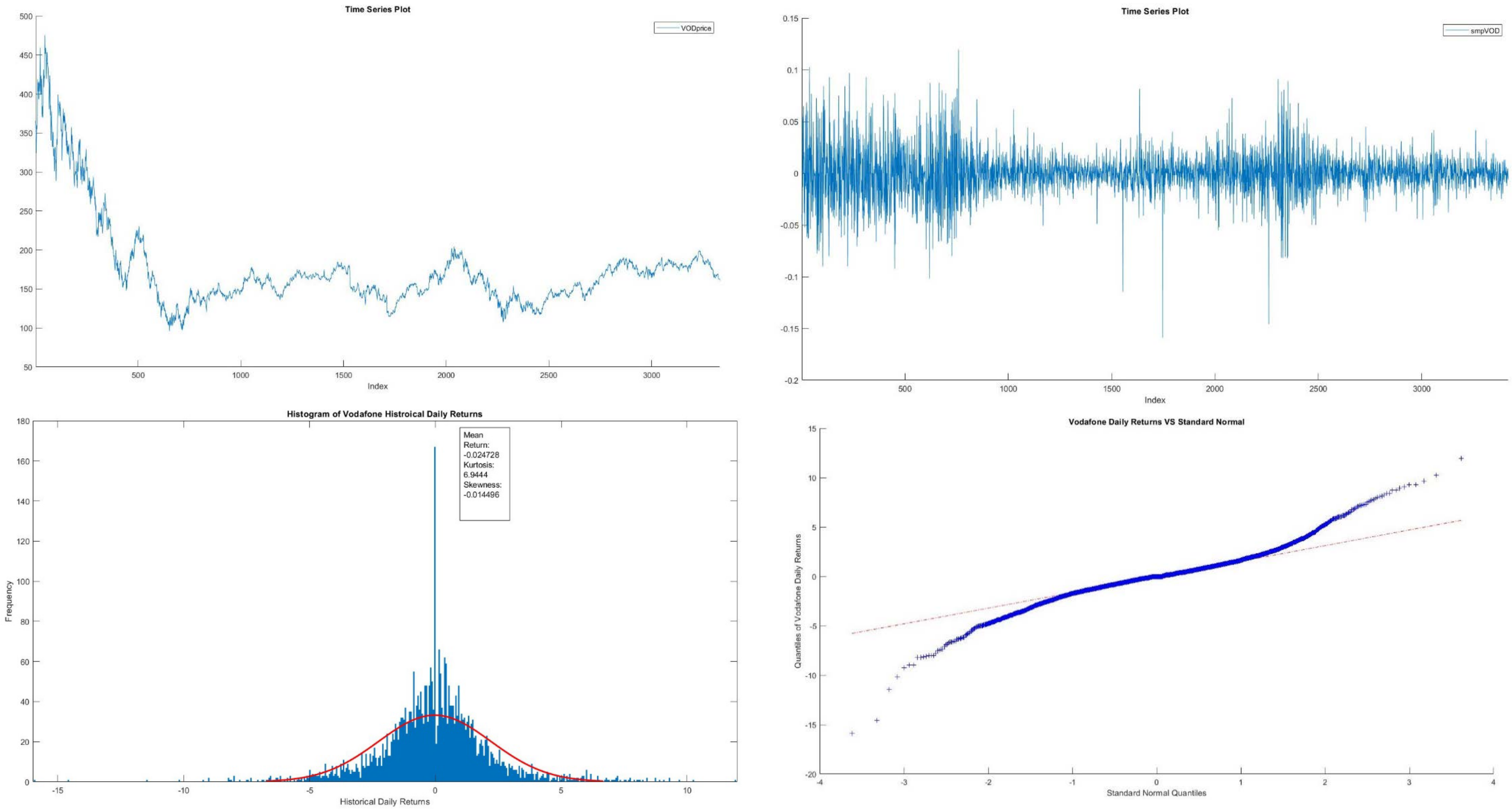


Figure 54: Historical Movement of Vodafone's Daily Closing Prices, Daily Returns, Histogram of Daily Returns and QQ Plot of Daily Returns and Standard Normal Quantiles

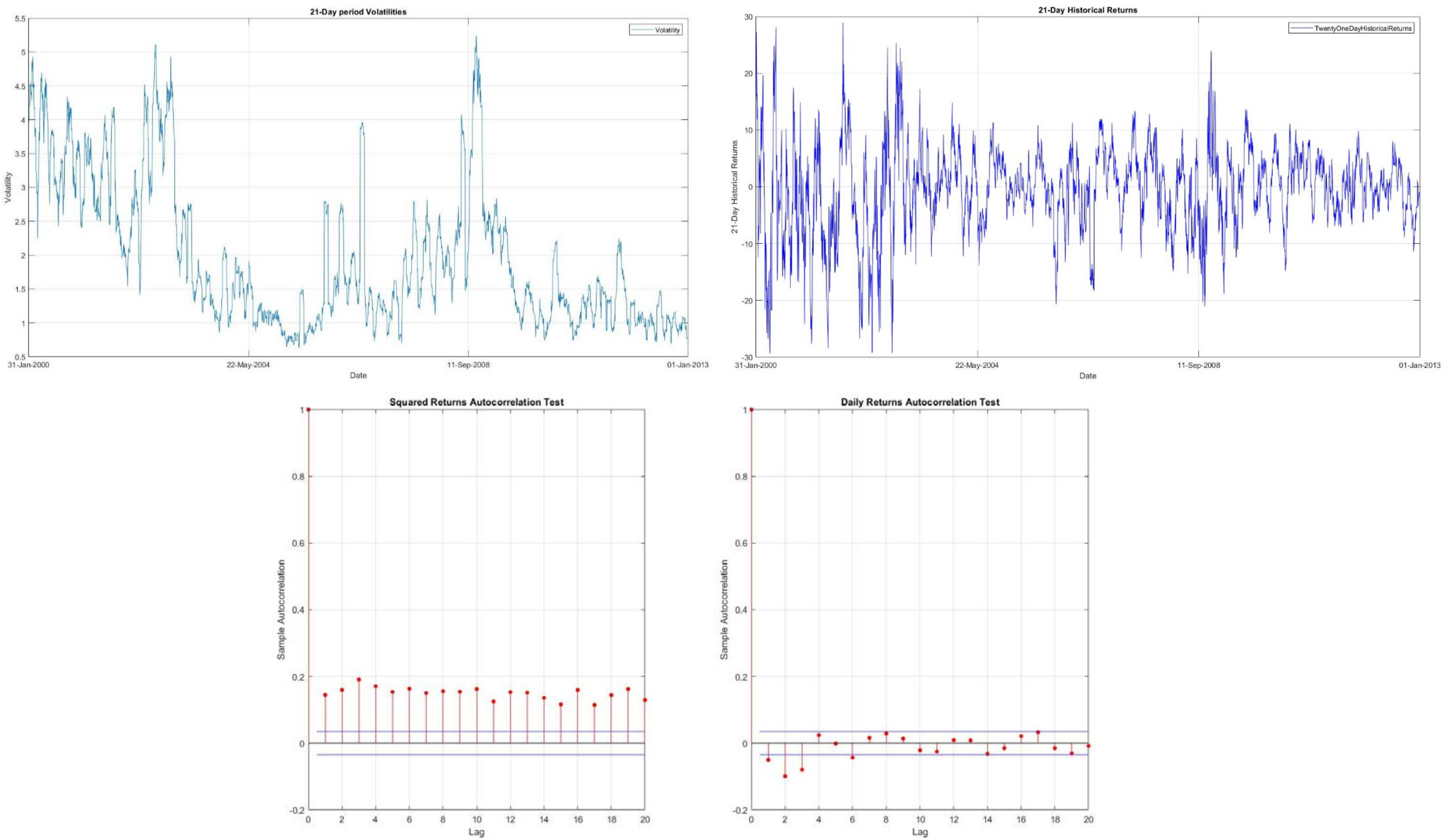


Figure 55: Historical Movement of Vodafone's 21-day Volatility, 21-day Returns and Autocorrelation for Squared Returns and Returns

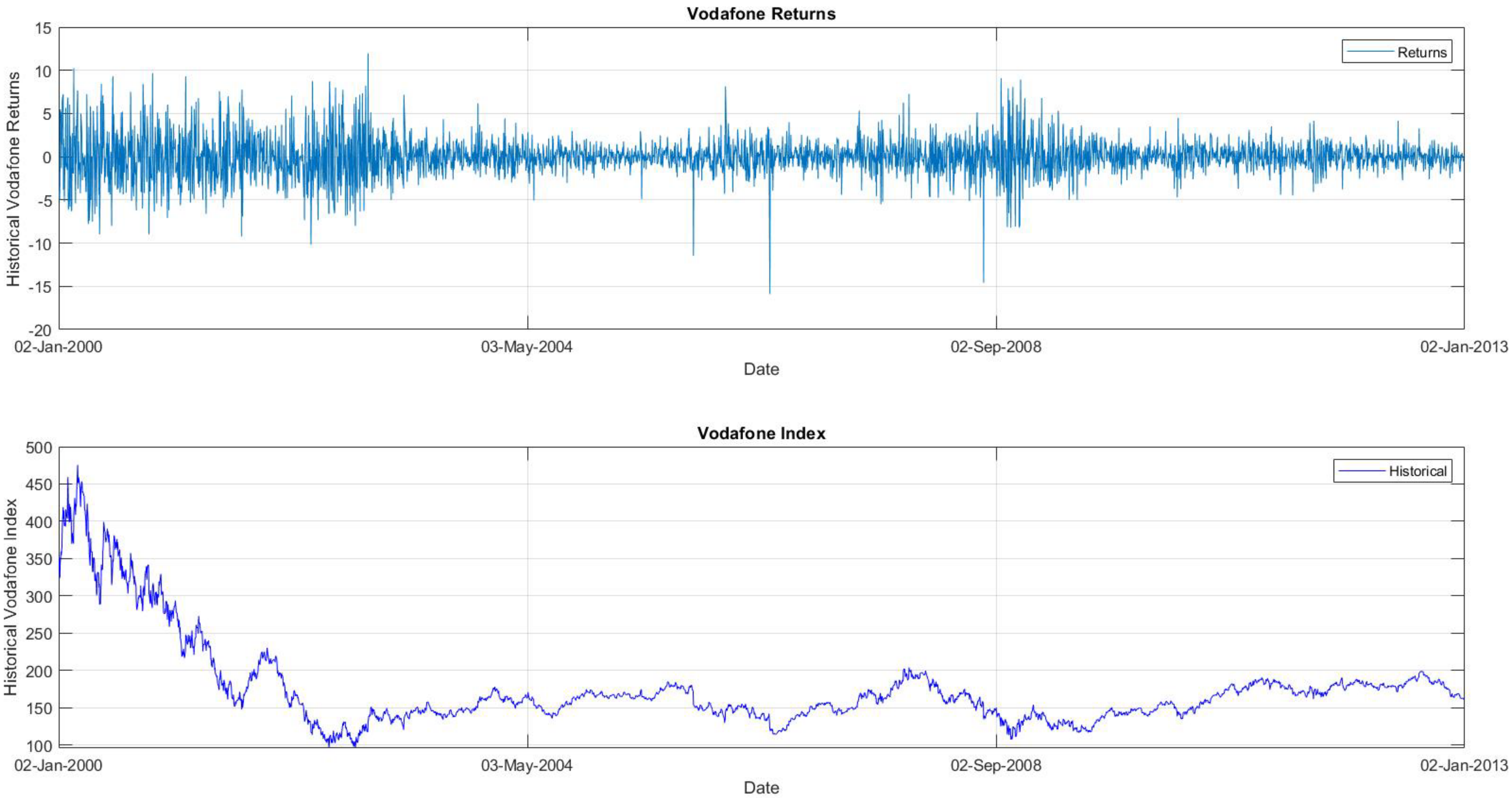


Figure 56: Comparison between Vodafone's Closing Daily Prices and Historical Returns

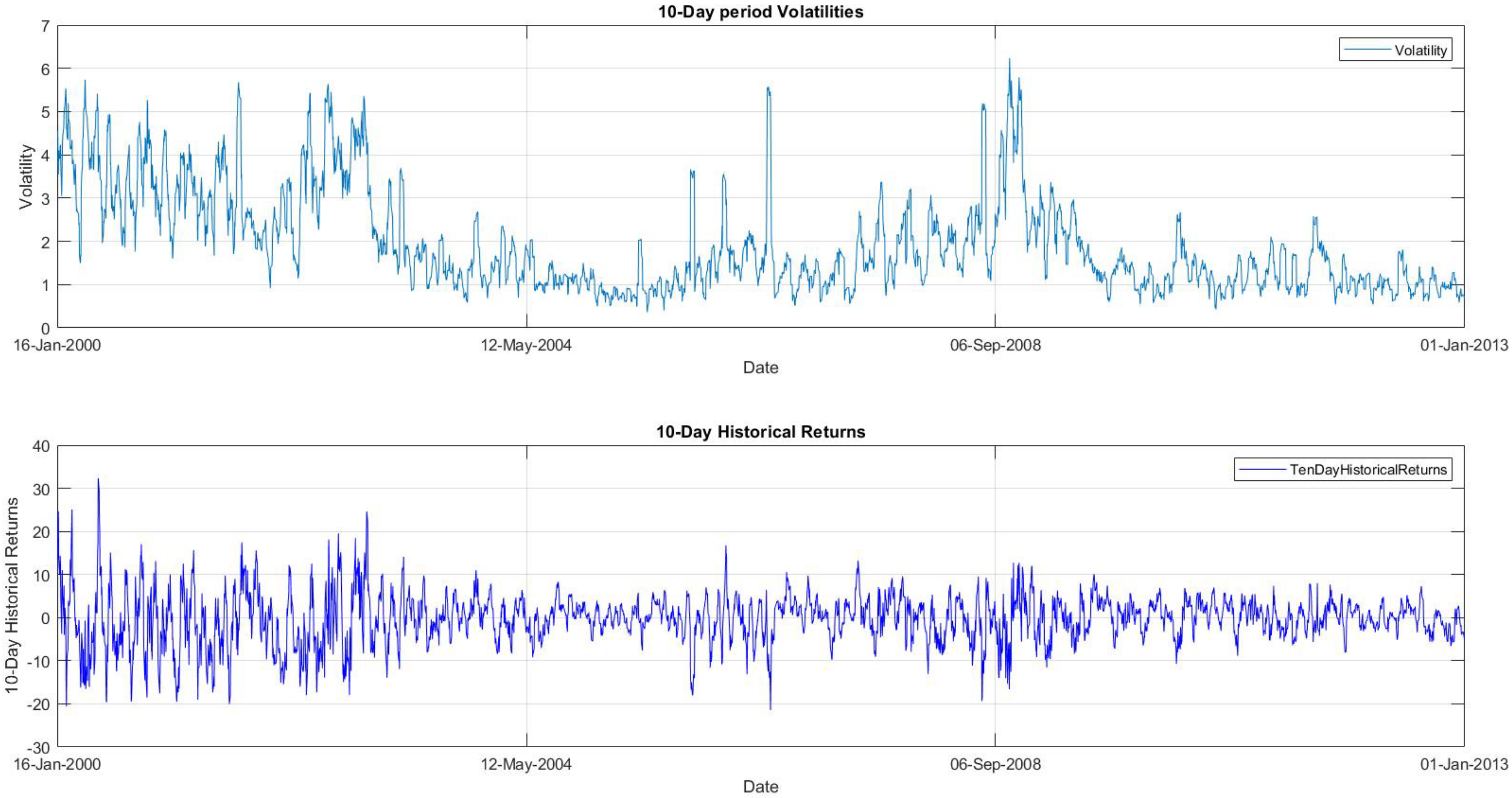


Figure 57: Comparison between Vodafone's 10-Day Period Volatilities and 10 Day Historical Returns

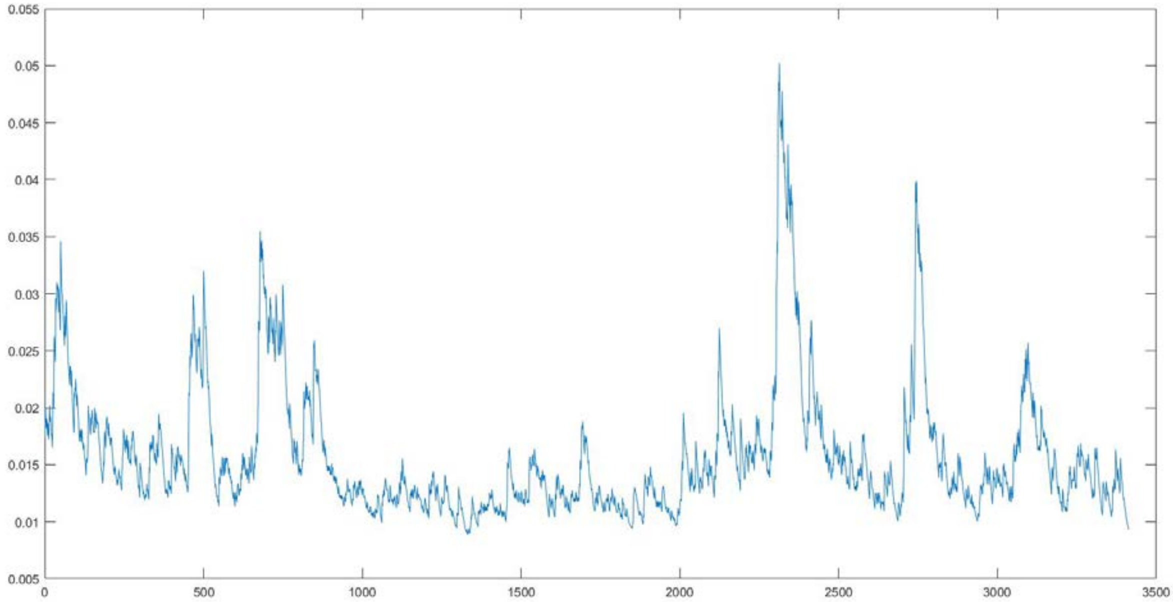


Figure 58: Model 1 - Volatility forecast of BP returns

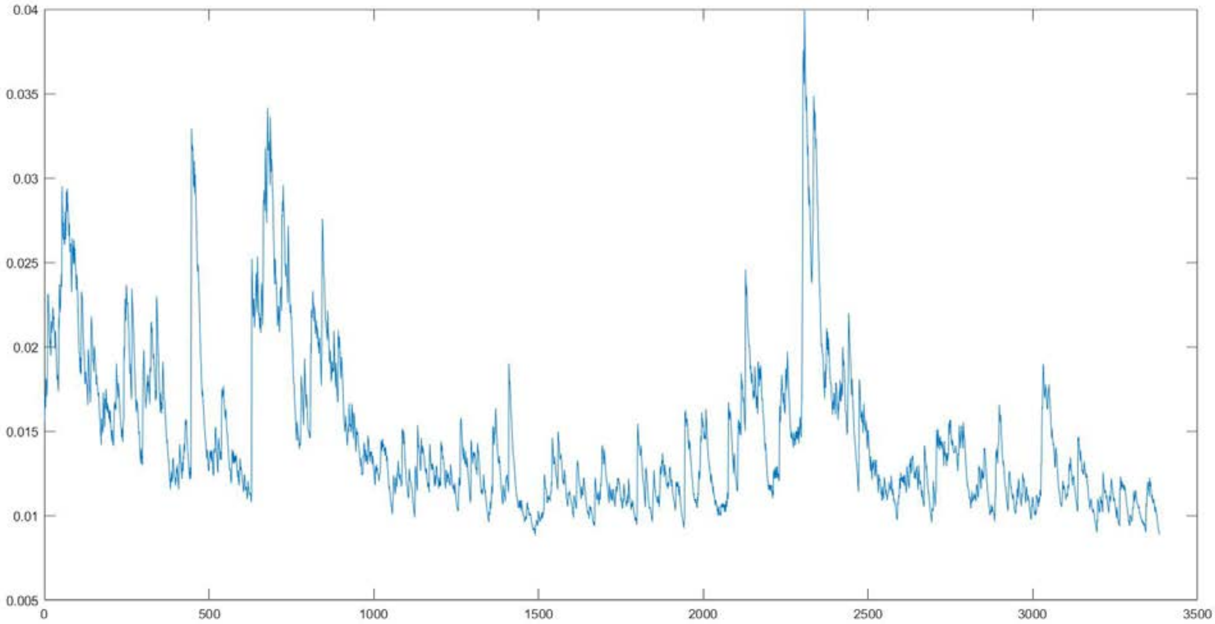


Figure 59: Model 1 - Volatility forecast of GlaxoSmithKline returns

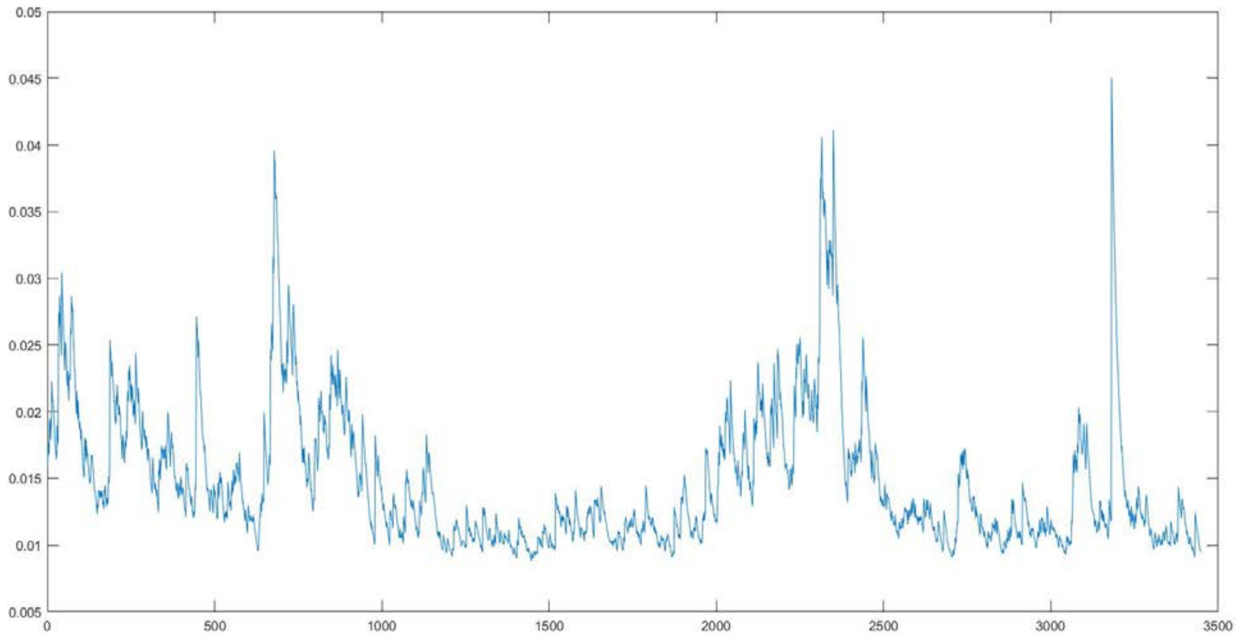


Figure 60: Model 1 - Volatility forecast of Tesco returns

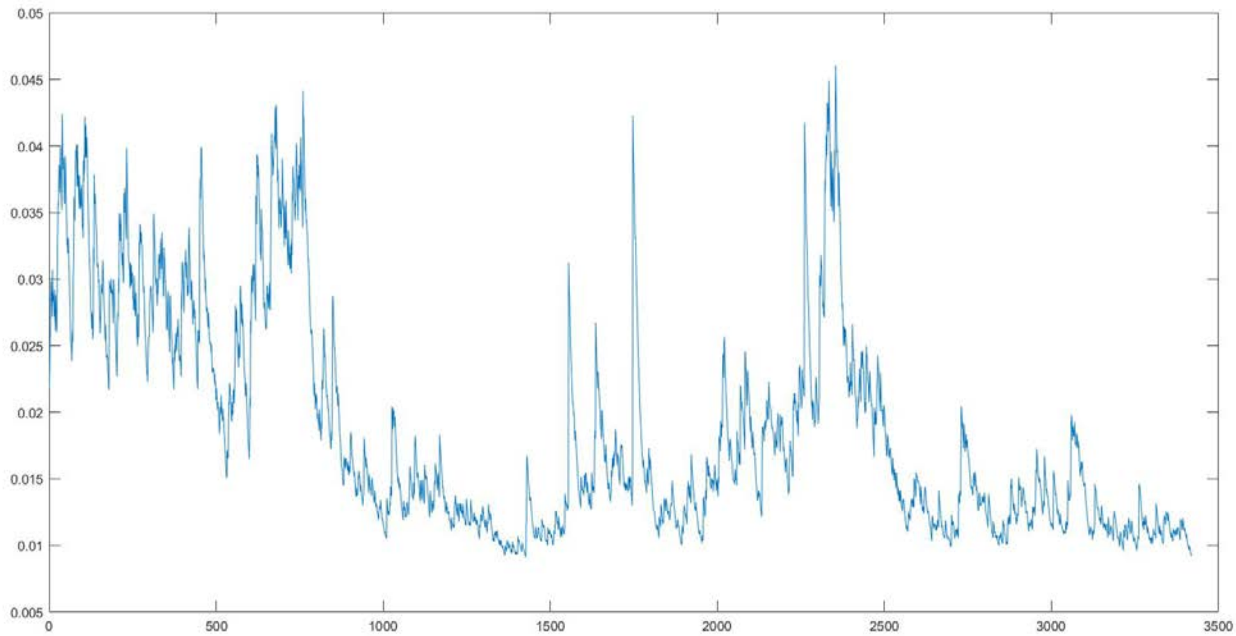


Figure 61: Model 1 - Volatility forecast of Vodafone returns

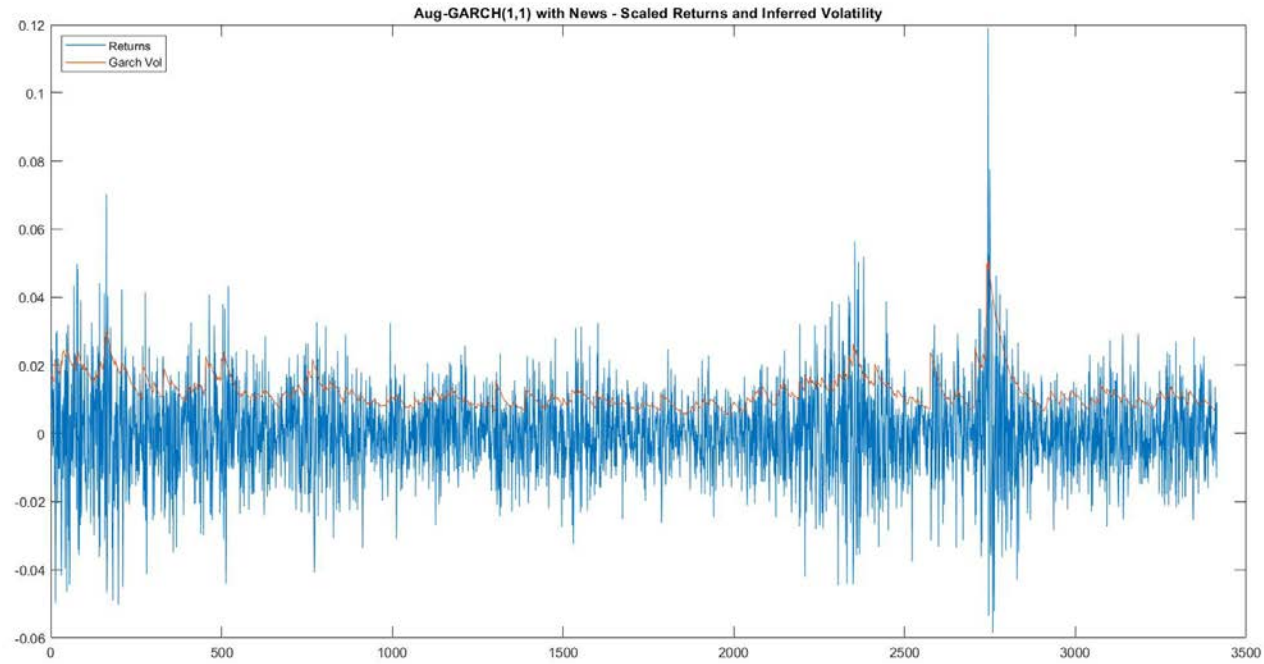


Figure 62: Inferred Volatility forecast and Scaled Returns of Aug-GARCH(1,1) with News model for BP

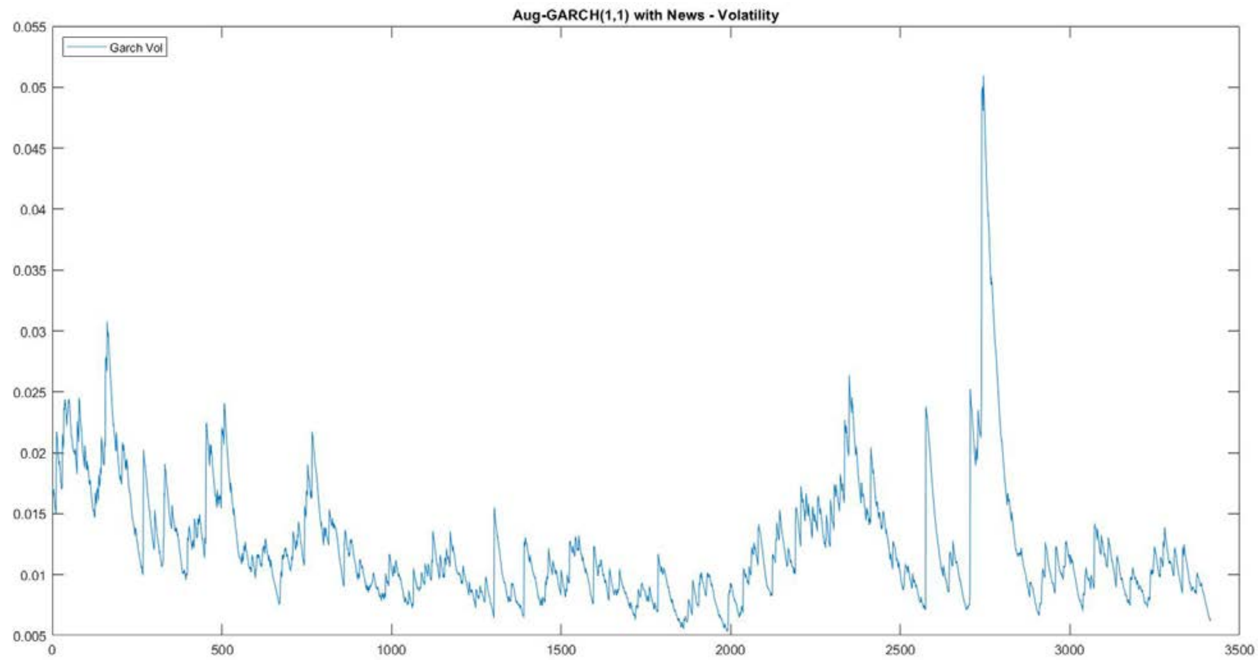


Figure 63: Volatility forecast of Aug-GARCH(1,1) with News model for BP returns

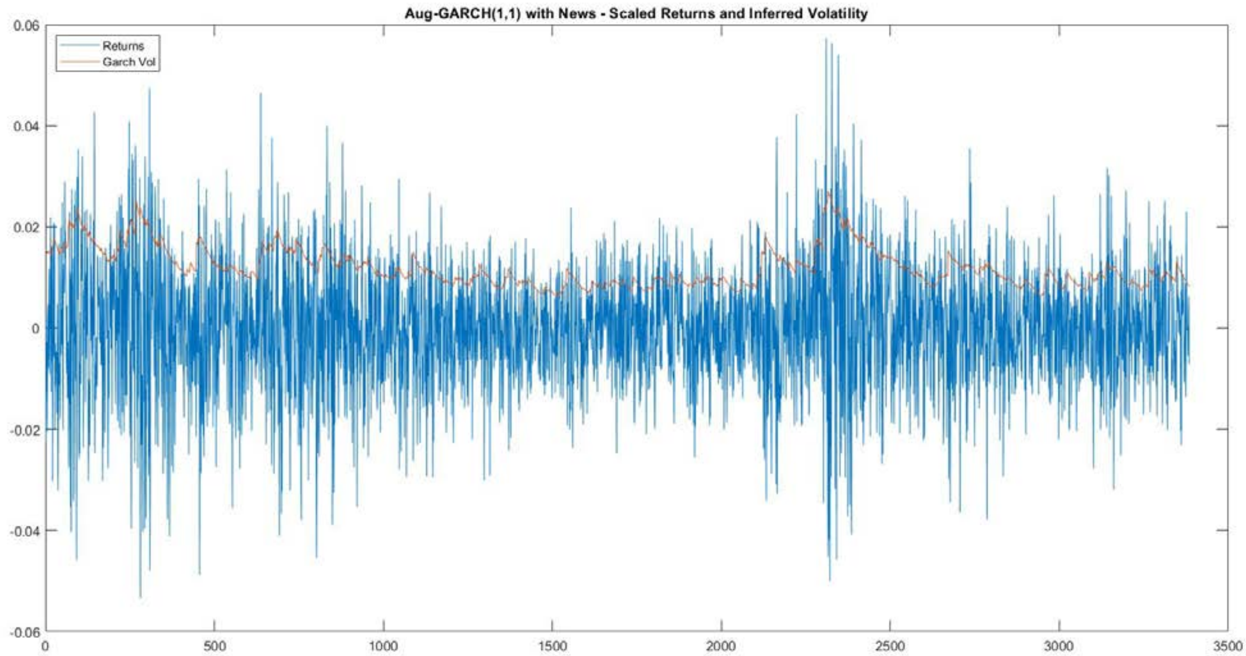


Figure 64: Inferred Volatility forecast and Scaled Returns of Aug-GARCH(1,1) with News model for GlaxoSmithKline

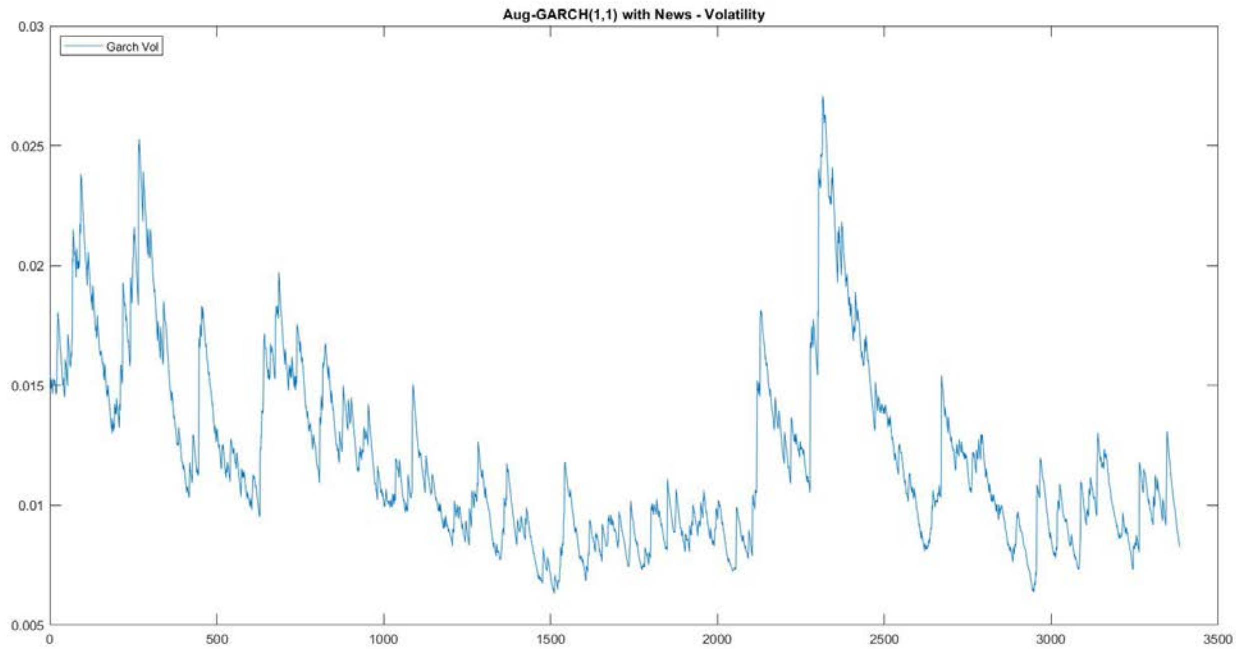


Figure 65: Volatility forecast of Aug-GARCH(1,1) with News model for GlaxoSmithKline returns

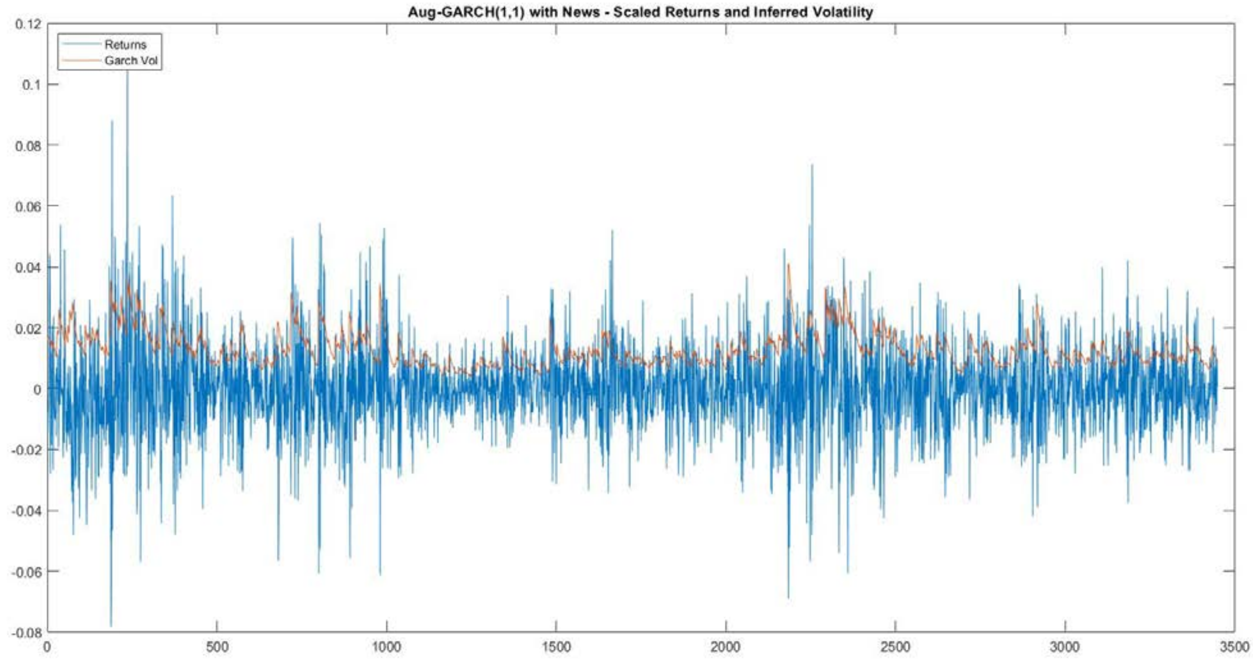


Figure 66: Inferred Volatility forecast and Scaled Returns of Aug-GARCH(1,1) with News model for Tesco

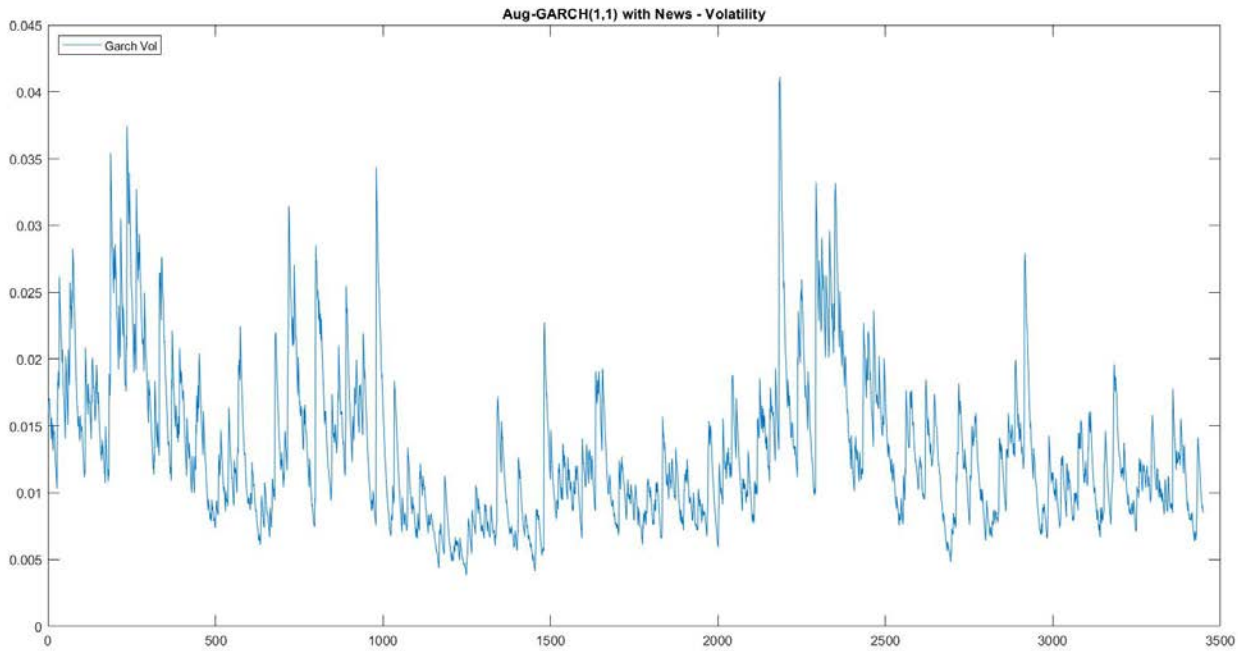


Figure 67: Volatility forecast of Aug-GARCH(1,1) with News model for Tesco returns

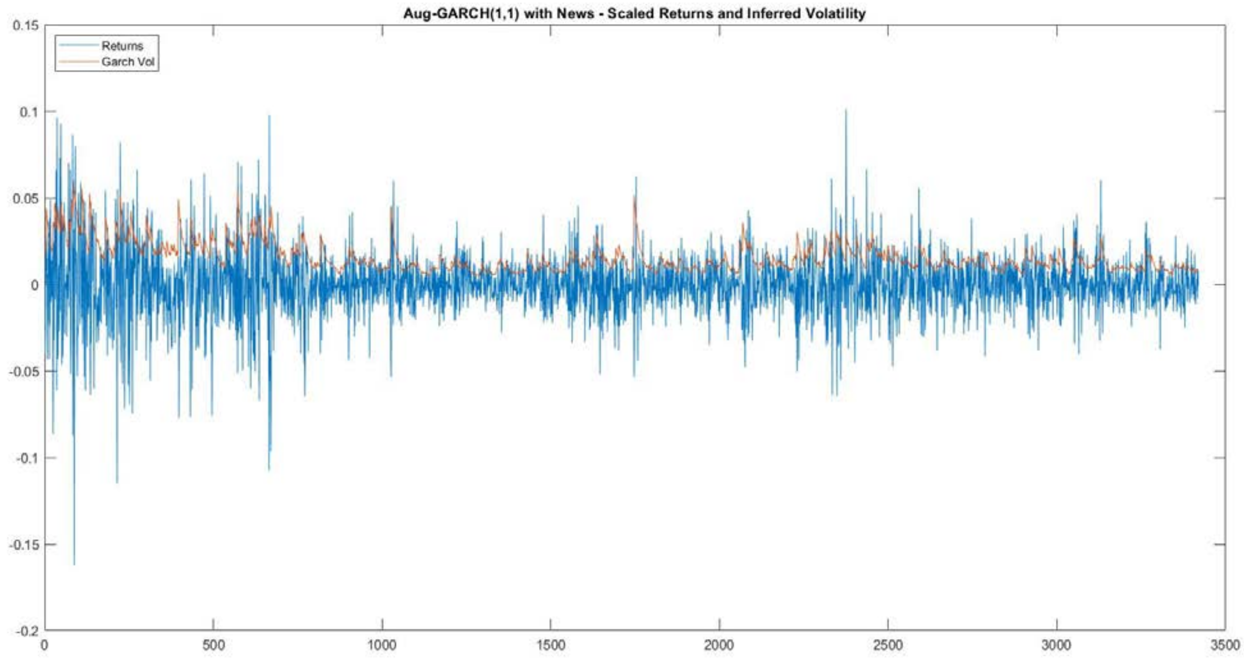


Figure 68: Inferred Volatility forecast and Scaled Returns of Aug-GARCH(1,1) with News model for Vodafone

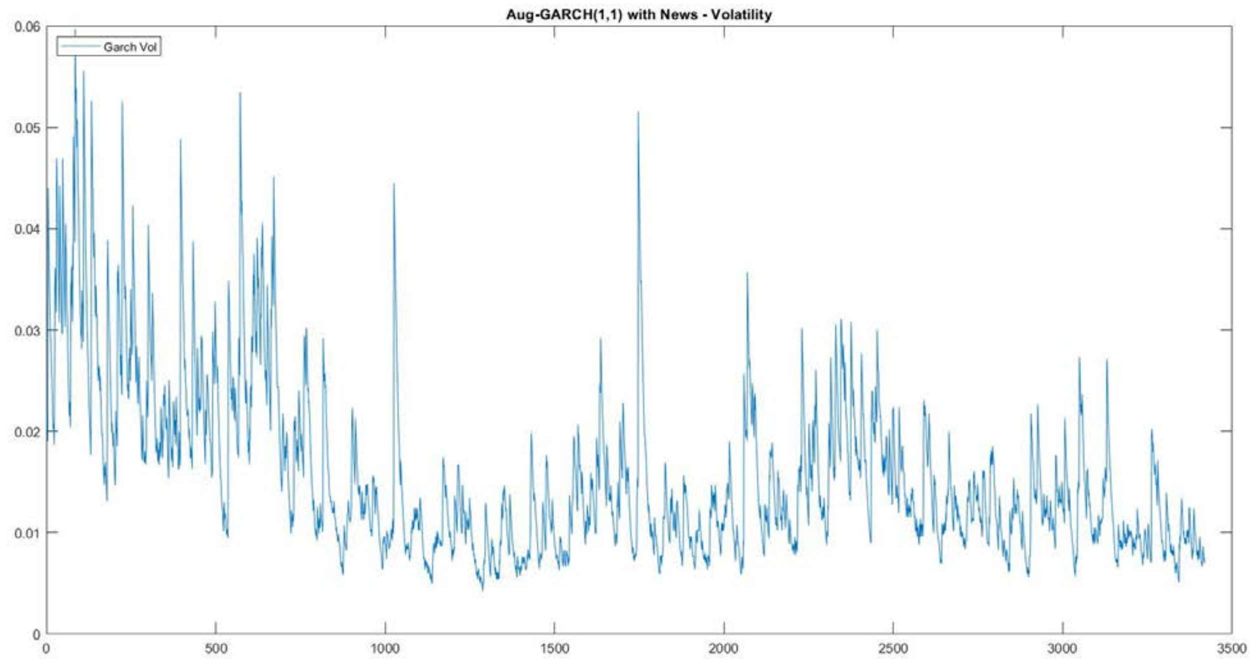


Figure 69: Volatility forecast of Aug-GARCH(1,1) with News model for Vodafone returns

Appendix C – Tables

Average Tone Values of Total Releases, Media Articles and Press Releases per Day of the Week

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Total Releases	0.3477	0.3605811	0.3459751	0.342111	0.293648	0.243156	0.237031
Media Articles	0.28228	0.3085321	0.2871898	0.28205	0.237946	0.243068	0.234035
Press Releases	0.40375	0.4160173	0.4162723	0.411017	0.377201	0.24734	0.290827

Average Tone Values of Total Releases, Media Articles and Press Releases per Month of the Year

	January	February	March	April	May	June	July	August	September	October	November	December
Total Releases	0.33235	0.3154426	0.36356	0.34166	0.342862	0.3565788	0.326282	0.31005	0.353963	0.309825	0.3305446	0.3618099
Media Articles	0.2864	0.2698863	0.30567	0.28203	0.2747013	0.3030476	0.26253	0.265027	0.277876	0.253447	0.2750118	0.2987854
Press Releases	0.43749	0.3784194	0.41587	0.40366	0.4060892	0.4057884	0.402261	0.363654	0.437271	0.40526	0.406961	0.4347857

Average Tone Values of Total Releases, Media Articles and Press Releases per Year

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Total Releases	0.4374869	0.33179	0.3026718	0.29467	0.35681	0.3502583	0.4269801	0.411176	0.322959	0.26384	0.338284	0.3017753	0.281718	0.40055	0.43085
Media Articles	0.3513632	0.20003	0.215749	0.23644	0.30419	0.3283706	0.38692	0.359108	0.232469	0.211284	0.323279	0.2601247	0.2293702	-	-
Press Releases	0.5242926	0.47713	0.401267	0.35473	0.41488	0.3730515	0.4652162	0.463906	0.416611	0.343508	0.36593	0.365579	0.3655317	0.40055	0.43085

Table 27: Average Volume of Total Releases, Media Articles and Press Releases by Calendar Period

Appendix D – Programme Code

Model 1: GARCH(1,1)

```
% GARCH_1_1_run.m

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Code: GARCH_1_1_run.m                        %
% Student Name: Jonathan Nolan                 %
% Student Number: 16071514                     %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%-----------------------------------------------
% Clear all work spaces and previous code
clc
clear
%-----------------------------------------------
% Load your data sets - Choose from 5 companies:
% 1. ASTRAZENECA
%COMP = xlsread('astraftseN.xlsx',1, 'A:G','basic');
% 2. BP
%COMP = xlsread('bpftseN.xlsx',1, 'A:G','basic');
% 3. GLAXOSMITHKLINE
%COMP = xlsread('glaxoftseN.xlsx',1, 'A:G','basic');
% 4. TESCO
%COMP = xlsread('tescoftseN.xlsx',1, 'A:G','basic');
% 5. VODAFONE
%COMP = xlsread('vodaftseN.xlsx',1, 'A:G','basic');

%-----------------------------------------------
% Calculate the log returns:
ret = diff(log(COMP(:,3)));
ret(isnan(ret))=0;
ret(~isfinite(ret))=0;
%-----------------------------------------------
% Initial Parameters
% NOTE: These will be used to find initial points of the search
Initial_1 = 9;
Initial_2 = 5;

%-----------------------------------------------
% Calculates opitmal params for function below
Params = GARCH_1_1_runALL(Initial_1,Initial_2, ret);

%-----------------------------------------------
% Log Likelihood Function
LLF = -GARCH_1_1_Maxlikelihood(ret, Params);
```

```
% GARCH_1_1_runALL.m
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Code: GARCH_1_1_runALL.m                     %
% Student Name: Jonathan Nolan                 %
% Student Number: 16071514                     %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function d = GARCH_1_1_runALL (Initial_1,Initial_2, ret)
%-----------------------------------------------
% Initial Values
Minimum_Value=0;
x = 9;
z = 0;

%-----------------------------------------------
for a=x:Initial_1
for b=z:(Initial_2 - 1)
for c=z:(Initial_2 - b - 1)
%-------------------------------------------------
% The parameters ininitial values:

parameter_1=10^(-a);
parameter_2=b/Initial_2+0.0001;
parameter_3=c/Initial_2+0.0001;

%-------------------------------------------------
% Put these in to a dataset
startParams = [parameter_1 parameter_2 parameter_3];

Parameters = GARCH_1_1_calibration ...
(ret, startParams);

G_val = GARCH_1_1_Maxlikelihood ...
(ret, Parameters);

if (G_val < Minimum_Value)
Minimum_Value=G_val;
d = Parameters;

end;

end;

end;

end;
```

```
% Aug_GARCH_1_1_calibration.m

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Code: GARCH_1_1_calibration.m                %
% Student Name: Jonathan Nolan                 %
% Student Number: 16071514                     %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function op_params = GARCH_1_1_calibration ...
(ret, StartParameters)

function h = mns_aux(Parameters)
h = GARCH_1_1_Maxlikelihood(ret, Parameters);

end

% Calculates opitmal params
op_params = fminsearch(@mns_aux, StartParameters);

end
```

```
% GARCH_1_1_Maxlikelihood.m

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Code: GARCH_1_1_Maxlikelihood.m              %
% Student Name: Jonathan Nolan                 %
% Student Number: 16071514                     %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%------------------------------------------------
function j = GARCH_1_1_Maxlikelihood ...
(returns, Params)
%------------------------------------------------
% The parameters initial values:

parameter_one=Params(1);

parameter_two=Params(2);

parameter_three=Params(3);
%------------------------------------------------
% the datasets length
num=length(returns);

if ((parameter_one<0) || (parameter_two<0) || (parameter_three<0))

j=intmax;
% Code returns the maximum integer

return;
end

variance_t(1,1)=nanvar(returns);
j = -log(variance_t(1))-(returns(1)^2/variance_t(1));

for count=2:num

variance_t(count,1)=parameter_one+parameter_two*(returns(count-1))^2+ ...
parameter_three*variance_t(count-1);

j=j-log(variance_t(count))-...
((returns(count))^2/variance_t(count));

end

j=-(1/2)*(j-log(2*pi));

end
```

Model 2: Aug-GARCH(1,1) model – Volume (V_t)

```
% Aug_GARCH_1_1_vol.m

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Code: Aug_GARCH_1_1_vol.m                    %
% Student Name: Jonathan Nolan                 %
% Student Number: 16071514                     %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%-----------------------------------------------
% Clear all work spaces and previous code

clc
clear

%-----------------------------------------------
% Initial Parameters
% NOTE: These will be used to find initial points of the search

Initial_1=9;
Initial_2=5;
Initial_3=5;

%-----------------------------------------------
% Load your data sets - Choose from 5 companies:

% 1. ASTRAZENECA
COMP = xlsread('astraftseN.xlsx',1, 'A:G','basic');
% 2. BP
%COMP = xlsread('bpftseN.xlsx',1, 'A:G','basic');
% 3. GLAXOSMITHKLINE
%COMP = xlsread('glaxoftseN.xlsx',1, 'A:G','basic');
% 4. TESCO
%COMP = xlsread('tescoftseN.xlsx',1, 'A:G','basic');
% 5. VODAFONE
%COMP = xlsread('vodaftseN.xlsx',1, 'A:G','basic');
%-----------------------------------------------
% Calculate the log returns:
ret = diff(log(COMP(:,3)));
ret(isnan(ret))=0;
ret(~isfinite(ret))=0;


%-----------------------------------------------
ftse_rets = diff(log(COMP(:,6)));
ftse_rets(isnan(ftse_rets))=0;
ftse_rets(~isfinite(ftse_rets))=0;
```

```
%-----------------------------------------------
% load the volume of news published per day
T = length(COMP);

%VOLUME
vol = COMP(2:T,4);
volume = vol/(max(vol));


Parameters = Aug_GARCH_1_1_runALL_vol(Initial_1,Initial_2,Initial_3,ret,volume);


LLF = -Aug_GARCH_Maxlikelihood_vol(ret, volume, Parameters);
```

```
% Aug_GARCH_1_1_runALL_vol.m

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Code: Aug_GARCH_1_1_runALL_vol.m        %
% Student Name: Jonathan Nolan                 %
% Student Number: 16071514                     %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function j = Aug_GARCH_1_1_runALL_vol (Initial_1,Initial_2,Initial_3, ret,
Trad_vol)
%-----------------------------------------------
% Initial Values
Minimum_Value=0;
x = 9;
z = 0;
y = 4;

for a=x:Initial_1
for b=z:(Initial_2-1)
for c=z:(Initial_2-b-1)
for d=y:Initial_3
%-------------------------------------------------
% The parameters ininitial values:

parameter_1=10^(-a);
parameter_2=b/Initial_2+0.0001;
parameter_3=c/Initial_2+0.0001;
parameter_4=10^(-d);
%-------------------------------------------------
% Put these in to a dataset
startParams = [parameter_1 parameter_2 parameter_3 parameter_4];

Parameters = Aug_GARCH_1_1_calibration_vol ...
(ret, Trad_vol, startParams);

G_val = Aug_GARCH_Maxlikelihood_vol ...
(ret, Trad_vol, Parameters);

if (G_val<Minimum_Value)
Minimum_Value=G_val;
j = Parameters;
end;
end;
end;
end;
end;
```

```
% Aug_GARCH_1_1_calibration_vol.m

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Code: Aug_GARCH_1_1_calibration_vol.m        %
% Student Name: Jonathan Nolan                 %
% Student Number: 16071514                     %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function params = Aug_GARCH_1_1_calibration_vol ...
(ret, trad_vol, start_Parameters)

function j = mns_aux(Parameters)

j = Aug_GARCH_Maxlikelihood_vol(ret, trad_vol, Parameters);

end

params = fminsearch(@mns_aux, start_Parameters);

end
```

```
% Aug_GARCH_Maxlikelihood_vol.m

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Code: Aug_GARCH_Maxlikelihood_vol.m          %
% Student Name: Jonathan Nolan                 %
% Student Number: 16071514                     %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%-------------------------------------------------
function j = Aug_GARCH_Maxlikelihood_vol ...
(returns, trad_vol, Params)
%-------------------------------------------------
% The parameters ininitial values:

parameter_one=Params(1);
parameter_two=Params(2);
parameter_three=Params(3);
parameter_four=Params(4);
%-------------------------------------------------
% the datasets length
num=length(returns);

if ((parameter_one<0) || (parameter_two<0) || (parameter_three<0) ||
(parameter_four<0))

j=intmax;
% Code returns the maximum integer

return;
end

variance_t(1,1)=nanvar(returns);

j = -log(variance_t(1))-(returns(1)^2/variance_t(1));

for count=2:num

variance_t(count,1)=parameter_one+parameter_two*(returns(count-1))^2+ ...
parameter_three*variance_t(count-1)+parameter_four*trad_vol(count);

j=j-log(variance_t(count))-...
((returns(count))^2/variance_t(count));

end
j=-(1/2)*(j-log(2*pi));
end
```

Model 3: Aug-GARCH(1,1) model - Lagged Volume (V_{t-1}).

```
% Aug_GARCH_1_1_vol_lag.m

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Code: Aug_GARCH_1_1_vol_lag.m                %
% Student Name: Jonathan Nolan                 %
% Student Number: 16071514                     %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%-----------------------------------------------
% Clear all work spaces and previous code
clc
clear
%-----------------------------------------------
% Initial Parameters
% NOTE: These will be used to find initial points of the search
Initial_1=9;
Initial_2=5;
Initial_3=5;
%-----------------------------------------------
% Load your data sets - Choose from 5 companies:
% 1. ASTRAZENECA
%COMP = xlsread('astraftseN.xlsx',1, 'A:G','basic');
% 2. BP
%COMP = xlsread('bpftseN.xlsx',1, 'A:G','basic');
% 3. GLAXOSMITHKLINE
%COMP = xlsread('glaxoftseN.xlsx',1, 'A:G','basic');
% 4. TESCO
%COMP = xlsread('tescoftseN.xlsx',1, 'A:G','basic');
% 5. VODAFONE
COMP = xlsread('vodaftseN.xlsx',1, 'A:G','basic');
%-----------------------------------------------
% Calculate the log returns:
ret = diff(log(COMP(:,3)));
ret(isnan(ret))=0;
ret(~isfinite(ret))=0;
%-----------------------------------------------
% load the volume of news published per day
T = length(COMP);
%VOLUME
vol = COMP(2:T,4);
volume = vol/(max(vol));

Op_Params=Aug_GARCH_1_1_runALL_vol_lag(Initial_1,Initial_2,Initial_3,ret,volume);

LLF = -Aug_GARCH_Maxlikelihood_vol_lag(ret, volume, Op_Params);
```

```
% Aug_GARCH_1_1_runALL_news_vol_lag.m

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Code: Aug_GARCH_1_1_runALL_news_vol_lag.m    %
% Student Name: Jonathan Nolan                 %
% Student Number: 16071514                     %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function j = Aug_GARCH_1_1_runALL_vol_lag (Initial_1,Initial_2,Initial_3, ret,
trad_vol)
%-----------------------------------------------
% Initial Values
Minimum_Value=0;
x = 9;
z = 0;
y = 4;

for a=x:Initial_1
for b=z:(Initial_2-1)
for c=z:(Initial_2-b-1)
for d=y:Initial_3
%-------------------------------------------------
% The parameters ininitial values:

parameter_one=10^(-a);
parameter_two=b/Initial_2+0.0001;
parameter_three=c/Initial_2+0.0001;
parameter_four=10^(-d);
%-------------------------------------------------
% Put these in to a dataset
startParams = [parameter_one parameter_two parameter_three parameter_four];

Parameters = Aug_GARCH_1_1_calibration_vol_lag ...
(ret, trad_vol, startParams);

G_val = Aug_GARCH_Maxlikelihood_vol_lag ...
(ret, trad_vol, Parameters);

if (G_val<Minimum_Value)
Minimum_Value=G_val;
j = Parameters;
end;
end;
end;
end;
end;
```

```
% Aug_GARCH_1_1_calibration_news_vol_lag.m

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Code: Aug_GARCH_1_1_calibration_vol_lag.m    %
% Student Name: Jonathan Nolan                 %
% Student Number: 16071514                     %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function params = Aug_GARCH_1_1_calibration_vol_lag ...
(returns, trad_vol, start_Parameters)

function j = mns_aux(Parameters)
j = Aug_GARCH_Maxlikelihood_vol_lag(returns, trad_vol, Parameters);

end

params = fminsearch(@mns_aux, start_Parameters);

end
```

```
% Aug_GARCH_Maxlikelihood_vol_lag.m

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Code: Aug_GARCH_Maxlikelihood_vol_lag.m      %
% Student Name: Jonathan Nolan                 %
% Student Number: 16071514                     %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%-------------------------------------------------
function j = Aug_GARCH_Maxlikelihood_vol_lag ...
(returns, trad_vol, Params)
%-------------------------------------------------
% The parameters ininitial values:

parameter_one=Params(1);
parameter_two=Params(2);
parameter_three=Params(3);
parameter_four=Params(4);
%-------------------------------------------------
% the datasets length
num=length(returns);

if ((parameter_one<0) || (parameter_two<0) || (parameter_three<0) ||
(parameter_four<0))

j=intmax;
% Code returns the maximum integer

return;
end

variance_t(1,1)=nanvar(returns);
j = -log(variance_t(1))-(returns(1)^2/variance_t(1));

for count=2:num

variance_t(count,1)=parameter_one+parameter_two*(returns(count-1))^2+ ...
parameter_three*variance_t(count-1)+parameter_four*trad_vol(count);

j=j-log(variance_t(count))-...
((returns(count))^2/variance_t(count-1));

end

j=-(1/2)*(j-log(2*pi));

end
```

Model 4: Aug-GARCH(1,1) model – News ($news_t$)

```
% Aug_GARCH_1_1_news.m

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Code: Aug_GARCH_1_1_news.m                   %
% Student Name: Jonathan Nolan                 %
% Student Number: 16071514                     %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%-----------------------------------------------
% Clear all work spaces and previous code

clc
clear

%-----------------------------------------------
% Initial Parameters
% NOTE: These will be used to find initial points of the search

Initial_1=9;
Initial_2=5;
Initial_3=5;

%-----------------------------------------------
% Load your data sets - Choose from 5 companies:

% 1. ASTRAZENECA
COMP = xlsread('astraftseN.xlsx',1, 'A:G','basic');

% 2. BP
%COMP = xlsread('bpftseN.xlsx',1, 'A:G','basic');

% 3. GLAXOSMITHKLINE
%COMP = xlsread('glaxoftseN.xlsx',1, 'A:G','basic');

% 4. TESCO
%COMP = xlsread('tescoftseN.xlsx',1, 'A:G','basic');

% 5. VODAFONE
%COMP = xlsread('vodaftseN.xlsx',1, 'A:G','basic');

%-----------------------------------------------
% Calculate the log returns:
ret = diff(log(COMP(:,3)));
ret(isnan(ret))=0;
ret(~isfinite(ret))=0;

%-----------------------------------------------
```

```
ftse_rets = diff(log(COMP(:,6)));
ftse_rets(isnan(ftse_rets))=0;
ftse_rets(~isfinite(ftse_rets))=0;

%-----------------------------------------------
% load the volume of news published per day
T = length(COMP);
news_intensity = COMP(2:T,2);

Parameters=Aug_GARCH_1_1_runALL_news(Initial_1,Initial_2,Initial_3,ret,ftse_rets,ne
ws_intensity);

LLF = -Aug_GARCH_Maxlikelihood_news(ret,ftse_rets,news_intensity,Parameters);
```

```
% Aug_GARCH_1_1_runALL_news.m

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Code: Aug_GARCH_1_1_runALL_news.m            %
% Student Name: Jonathan Nolan                 %
% Student Number: 16071514                     %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function j = Aug_GARCH_1_1_runALL_news (Initial_1,Initial_2,Initial_3, returns,
ftse_rets, newsvolume)
%-----------------------------------------------
% Initial Values
Minimum_Value=0;
x = 9;
z = 0;
y = 4;
for a=x:Initial_1
for b=z:(Initial_2-1)
for c=z:(Initial_2-b-1)
for d=y:Initial_3
%-------------------------------------------------
% The parameters ininitial values:
parameter_1=10^(-a);
parameter_2=b/Initial_2+0.0001;
parameter_3=c/Initial_2+0.0001;
parameter_4=10^(-d);
parameter_5=0.0001;
parameter_6=1.0;
%-------------------------------------------------
% Put these in to a dataset
startParams = [parameter_1 parameter_2 parameter_3 parameter_4 parameter_5
parameter_6];

Parameters = Aug_GARCH_1_1_calibration_news ...
(returns, ftse_rets, newsvolume, startParams);
G_val = Aug_GARCH_Maxlikelihood_news ...
(returns, ftse_rets, newsvolume, Parameters);

if (G_val<Minimum_Value)
Minimum_Value=G_val;
j = Parameters;
end;
end;
end;
end;
end;
```

```
% Aug_GARCH_1_1_calibration_news.m

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Code: Aug_GARCH_1_1_calibration_news.m       %
% Student Name: Jonathan Nolan                 %
% Student Number: 16071514                     %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function parameters = Aug_GARCH_1_1_calibration_news ...
(returns, fts_rets, news_volume, start_Parameters)

function g = mns_aux(Parameters)
g = Aug_GARCH_Maxlikelihood_news(returns, fts_rets, news_volume, Parameters);

end;

parameters = fminsearch(@mns_aux, start_Parameters);

end;
```

```
% Aug_GARCH_Maxlikelihood_news.m

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Code: Aug_GARCH_Maxlikelihood_news.m         %
% Student Name: Jonathan Nolan                 %
% Student Number: 16071514                     %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%-------------------------------------------------
function j = Aug_GARCH_Maxlikelihood_news ...
(returns, ftse_rets, newsvolume, Parameters)

%-------------------------------------------------
% The parameters ininitial values:

parameter_one=Parameters(1);

parameter_two=Parameters(2);

parameter_three=Parameters(3);

parameter_four=Parameters(4);

parameter_five = Parameters(5);

parameter_six = Parameters(6);

%-------------------------------------------------
% the datasets length
num=length(returns);

if ((parameter_one<0) || (parameter_two<0) || (parameter_three<0) ||
(parameter_four<0))

j=intmax;
% Code returns the maximum integer

return;
end

variance_t(1,1)=nanvar(returns);
j = -log(variance_t(1))-(returns(1)^2/variance_t(1));

for count=2:num

variance_t(count,1)=parameter_one+parameter_two*(returns(count-1)-...
parameter_five-parameter_six*ftse_rets(count-1))^2+ ...
parameter_three*variance_t(count-1)+parameter_four*newsvolume(count);
```

```
j=j-log(variance_t(count))-...
((returns(count)-parameter_five-
parameter_six*ftse_rets(count))^2/variance_t(count));

end

j=-(1/2)*(j-log(2*pi));

end;
```

Model 5: Aug-GARCH(1,1) model - Lagged News $(news_{t-1})$

```
% Aug_GARCH_1_1_news_lag.m

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Code: Aug_GARCH_1_1_news_lag.m               %
% Student Name: Jonathan Nolan                 %
% Student Number: 16071514                     %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%-----------------------------------------------
% Clear all work spaces and previous code

clc
clear

%-----------------------------------------------
% Initial Parameters
% NOTE: These will be used to find initial points of the search

Initial_1=9;
Initial_2=5;
Initial_3=5;

%-----------------------------------------------
% Load your data sets - Choose from 5 companies:

% 1. ASTRAZENECA
COMP = xlsread('astraftseN.xlsx',1, 'A:G','basic');

% 2. BP
%COMP = xlsread('bpftseN.xlsx',1, 'A:G','basic');

% 3. GLAXOSMITHKLINE
%COMP = xlsread('glaxoftseN.xlsx',1, 'A:G','basic');

% 4. TESCO
%COMP = xlsread('tescoftseN.xlsx',1, 'A:G','basic');

% 5. VODAFONE
%COMP = xlsread('vodaftseN.xlsx',1, 'A:G','basic');

%-----------------------------------------------
% Calculate the log returns:
ret = diff(log(COMP(:,3)));
ret(isnan(ret))=0;
ret(~isfinite(ret))=0;
```

```
%-----------------------------------------------

fts_rets = diff(log(COMP(:,6)));
fts_rets(isnan(fts_rets))=0;
fts_rets(~isfinite(fts_rets))=0;

%-----------------------------------------------
% load the volume of news published per day
T = length(COMP);
news_intensity = COMP(2:T,2);


Parameters=Aug_GARCH_1_1_runALL_news_lag(Initial_1,Initial_2,Initial_3,ret,fts_rets
,news_intensity);


ANS = -Aug_GARCH_Maxlikelihood_news_lag(ret, fts_rets, news_intensity, Parameters)
```

```
% Aug_GARCH_1_1_runALL_news_lag.m

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Code: Aug_GARCH_1_1_runALL_news_lag.m        %
% Student Name: Jonathan Nolan                 %
% Student Number: 16071514                     %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%


function j = Aug_GARCH_1_1_runALL_news_lag (Initial_1,Initial_2,Initial_3, ret,
fts_rets, newsvolume)

%-----------------------------------------------
% Initial Values
Minimum_Value=0;
x = 9;
z = 0;
y = 4;

for a=x:Initial_1
for b=z:(Initial_2-1)
for c=z:(Initial_2-b-1)
for d=y:Initial_3


%-------------------------------------------------
% The parameters ininitial values:

%Omega
parameter_1=10^(-a);

%Alpha
parameter_2=b/Initial_2+0.0001;

%Beta
parameter_3=c/Initial_2+0.0001;

%Gamma
parameter_4=10^(-d);

%Theta1
parameter_5=0.0001;

%Theta2
parameter_6=1.0;

%-------------------------------------------------
% Put these in to a dataset
```

```
startParams = [parameter_1 parameter_2 parameter_3 parameter_4 parameter_5
parameter_6];

Parameters = Aug_GARCH_1_1_calibration_news_lag ...
(ret, fts_rets, newsvolume, startParams);

G_val = Aug_GARCH_Maxlikelihood_news_lag ...
(ret, fts_rets, newsvolume, Parameters);

if (G_val<Minimum_Value)
Minimum_Value=G_val;
j = Parameters;

end;
end;
end;
end;
end;
```

```
% Aug_GARCH_1_1_calibration_news_lag.m

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Code: Aug_GARCH_1_1_calibration_news_lag.m   %
% Student Name: Jonathan Nolan                 %
% Student Number: 16071514                     %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function parameters = Aug_GARCH_1_1_calibration_news_lag ...
(returns, fts_rets, newsvolume, start_Parameters)

function g = mns_aux(Parameters)
g = Aug_GARCH_Maxlikelihood_news_lag(returns, fts_rets, newsvolume, Parameters);

end

params = fminsearch(@mns_aux, start_Parameters);

end;
```

```
% Aug_GARCH_Maxlikelihood_news_lag.m

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Code: Aug_GARCH_Maxlikelihood_news_lag.m     %
% Student Name: Jonathan Nolan                 %
% Student Number: 16071514                     %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%-----------------------------------------------
function j = Aug_GARCH_Maxlikelihood_news_lag ...
(returns, fts_rets, newsvolume, Parameters)
%-----------------------------------------------
% The parameters ininitial values:
parameter_one=Parameters(1);
parameter_two=Parameters(2);
parameter_three=Parameters(3);
parameter_four=Parameters(4);
parameter_five = Parameters(5);
parameter_six = Parameters(6);
%-----------------------------------------------
% datasets length
num=length(returns);

if ((parameter_one<0) || (parameter_two<0) || (parameter_three<0) ||
(parameter_four<0))

j=intmax;
% Code returns the maximum integer
return;
end

variance_t(1,1)=nanvar(returns);
j = -log(variance_t(1))-(returns(1)^2/variance_t(1));

for count=2:num

variance_t(count,1)=parameter_one+parameter_two*(returns(count-1)-...
parameter_five-parameter_six*fts_rets(count-1))^2+ ...
parameter_three*variance_t(count-1)+parameter_four*newsvolume(count-1);
j=j-log(variance_t(count))-...
((returns(count)-parameter_five-
parameter_six*fts_rets(count))^2/variance_t(count));
end;
j=-(1/2)*(j-log(2*pi));

end;
```

Monte Carlo Simulations - Model 2: Aug-GARCH(1,1) model - Volume (V_t)

```
% Aug_GARCH_1_1_MonteCarlo_volume.m

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Code: Aug_GARCH_1_1_MonteCarlo_volume.m      %
% Student Name: Jonathan Nolan                 %
% Student Number: 16071514                     %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%-----------------------------------------------
% Clear all work spaces and previous code

clc
clear

%-----------------------------------------------
% Load your data sets - Choose from 5 companies:

% 1. ASTRAZENECA
COMP = xlsread('astraftseN.xlsx',1, 'A:G','basic');
% 2. BP
%COMP = xlsread('bpftseN.xlsx',1, 'A:G','basic');
% 3. GLAXOSMITHKLINE
%COMP = xlsread('glaxoftseN.xlsx',1, 'A:G','basic');
% 4. TESCO
%COMP = xlsread('tescoftseN.xlsx',1, 'A:G','basic');
% 5. VODAFONE
%COMP = xlsread('vodaftseN.xlsx',1, 'A:G','basic');

%-----------------------------------------------
% Calculate the log returns:
ret = diff(log(COMP(:,3)));
ret(isnan(ret))=0;
ret(~isfinite(ret))=0;

%-----------------------------------------------
% Initial Parameters
% NOTE: These will be used to find initial points of the search

Initial_1=9;
Initial_2=5;
Initial_3=5;

%---------------------------------------------------------
% load input arrays of trading volume per day
%---------------------------------------------------------
```

```
%Length of dataset
J = length(COMP);

%VOLUME
vol = COMP(2:J,4);
trade_vol = vol/(max(vol));

%---------------------------------------------------------
% Aug-GARCH(1,1) Parameters
%---------------------------------------------------------

% maximizing likelihood function
Parameters=Aug_GARCH_1_1_runALL_vol(Initial_1,Initial_2,Initial_3,ret,trade_vol);

%---------------------------------------------------------
% GARCH(1,1) Parameters
%---------------------------------------------------------

GARCH_1_1_Params = GARCH_1_1_runALL(Initial_1,Initial_2, ret);

%---------------------------------------------------------
% Other Variables
%---------------------------------------------------------

% 1. Time Period Length
Time = length(ret);

% 2. amount of simulations
Num=1000;

for x=1:Num

h=MC_GARCH_Volume_simulation ( ret, trade_vol, Time, Parameters);

for u=1:Time

ec(u) = h(u)*random('normal',0,1);

end
%---------------------------------------------------------
updated_rets=ec;
%---------------------------------------------------------
% Aug-GARCH(1,1)- VOLUME
%---------------------------------------------------------

Param_Aug_G_1_1_Vol =...
Aug_GARCH_1_1_calibration_vol (updated_rets, trade_vol, Parameters);
```

```
Parameter_One(x)=Param_Aug_G_1_1_Vol(1);
Parameter_Two(x)=Param_Aug_G_1_1_Vol(2);
Parameter_Three(x)=Param_Aug_G_1_1_Vol(3);
Parameter_Four(x)=Param_Aug_G_1_1_Vol(4);

LogLF_G_Vol(x)=...
-Aug_GARCH_Maxlikelihood_vol(updated_rets, trade_vol, Param_Aug_G_1_1_Vol);

%--------------------------------------------------------
% GARCH(1,1)
%--------------------------------------------------------
Parameters_GARCH_1_1 = GARCH_1_1_calibration(updated_rets, GARCH_1_1_Params);

LogLF_G(x)=...
-GARCH_1_1_Maxlikelihood(updated_rets, Parameters_GARCH_1_1);

%--------------------------------------------------------
% Likelihood Ratio
%--------------------------------------------------------
Log_L_Rat(x)=2*(LogLF_G_Vol(x)-LogLF_G(x));

end

%--------------------------------------------------------
% Calculates the average values of the parameters

average_parameter_one = mean(Parameter_One); %omega
average_parameter_two = mean(Parameter_Two); %alpha
average_parameter_three = mean(Parameter_Three); %beta
average_parameter_four = mean(Parameter_Four); %gamma

average_LoglikelihoodFunc=mean(LogLF_G_Vol);

%--------------------------------------------------------
% Calculates the variances of the parameters
var_parameter_one=var(Parameter_One);
var_parameter_two=var(Parameter_Two);
var_parameter_three=var(Parameter_Three);
var_parameter_four=var(Parameter_Four);
var_LogLF=var(LogLF_G_Vol);

%--------------------------------------------------------
% Histograms
%--------------------------------------------------------

% Histogram of Alpha
figure(1);
```

```
hist(Parameter_Two,30);
title('Histogram of Alpha');

% Histogram of Beta
figure(2);
hist(Parameter_Three,30);
title('Histogram of Beta')

% Histogram of Gamma
figure(3);
hist(Parameter_Four,30);
title('Histogram of Gamma')

%-------------------------------------------------------
% Plot for the Returns and Inferred Volatility
%-------------------------------------------------------

%NOTE: Appropriate values need to be arranged first:

% 1. Forcasted Returns
smp_ret (1:Time,1) = updated_rets(1,1:Time);

% 2. Scaled Forecasted Returns
smp_ret_mean = smp_ret - mean(smp_ret);

% 3. Forecasted Volatility
w_var (1:Time,1) = h(1,1:Time);
nonzeros(w_var);

%------------------------------------------------------
% Plot the figure
figure(7);
plot(smp_ret_mean); hold on
plot(w_var); hold off
title('Aug-GARCH(1,1) with Volume - Scaled Returns and Inferred Volatility')
legend('Returns', 'Aug-Garch(1,1) vol - Volatility', 'Location', 'northwest')

%------------------------------------------------------
% Plot the figure
figure(8);
plot(w_var); hold off
title('Aug-GARCH(1,1) with Volume - Volatility')
legend('Aug-Garch(1,1) vol - Volatility', 'Location', 'northwest')
%------------------------------------------------------
% Histogram of Gamma
figure(9);
hist(Log_L_Rat,30);
```

```
title('Histogram of Loglikelihood Ratio');
```

Monte Carlo Simulations - Model 3: Aug-GARCH(1,1) model – Lagged Volume (V_{t-1})

```
% Aug_GARCH_MonteCarlo_volume_lag.m

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Code: Aug_GARCH_MonteCarlo_volume_lag.m      %
% Student Name: Jonathan Nolan                 %
% Student Number: 16071514                     %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%-----------------------------------------------
% Clear all work spaces and previous code

clc
clear

%-----------------------------------------------
% Load your data sets - Choose from 5 companies:

% 1. ASTRAZENECA
COMP = xlsread('astraftseN.xlsx',1, 'A:G','basic');
% 2. BP
%COMP = xlsread('bpftseN.xlsx',1, 'A:G','basic');
% 3. GLAXOSMITHKLINE
%COMP = xlsread('glaxoftseN.xlsx',1, 'A:G','basic');
% 4. TESCO
%COMP = xlsread('tescoftseN.xlsx',1, 'A:G','basic');
% 5. VODAFONE
%COMP = xlsread('vodaftseN.xlsx',1, 'A:G','basic');

%-----------------------------------------------
% Calculate the log returns:
ret = diff(log(COMP(:,3)));
ret(isnan(ret))=0;
ret(~isfinite(ret))=0;

%-----------------------------------------------
% Initial Parameters
% NOTE: These will be used to find initial points of the search

Initial_1=9;
Initial_2=5;
Initial_3=5;
```

```
%--------------------------------------------------------
% load input arrays of trading volume per day
%--------------------------------------------------------

%Length of dataset
J = length(COMP);

%VOLUME
vol = COMP(2:J,4);
trade_vol = vol/(max(vol));

%--------------------------------------------------------
% Aug-GARCH(1,1) Lagged Vol Parameters
%--------------------------------------------------------

% maximizing likelihood function
Parameters=Aug_GARCH_1_1_runALL_vol_lag(Initial_1,Initial_2,Initial_3,ret,trade_vol
);
%--------------------------------------------------------
% GARCH(1,1) Parameters
%--------------------------------------------------------

GARCH_1_1_Params = GARCH_1_1_runALL(Initial_1,Initial_2, ret);
%--------------------------------------------------------
% Other Variables
%--------------------------------------------------------

Time = length(ret);
Num=1000;

for x=1:Num

w=Monte_Carlo_G_News_sim ( ret, trade_vol, Time, Parameters);

for u=1:Time
%--------------------------------------------------------
ec(u) = w(u)*random('normal',0,1);
end
%--------------------------------------------------------
updated_rets=ec;

%--------------------------------------------------------
% Aug-GARCH(1,1)- VOLUME Lagged
%--------------------------------------------------------
Param_Aug_G_Lag_Vol =...
Aug_GARCH_1_1_calibration_vol_lag (updated_rets, trade_vol, Parameters);
```

```
Parameter_One(x)=Param_Aug_G_Lag_Vol(1);
Parameter_Two(x)=Param_Aug_G_Lag_Vol(2);
Parameter_Three(x)=Param_Aug_G_Lag_Vol(3);
Parameter_Four(x)=Param_Aug_G_Lag_Vol(4);

LogLF_G_Lag_Vol(x)=...
-Aug_GARCH_Maxlikelihood_vol_lag(updated_rets, trade_vol, Param_Aug_G_Lag_Vol);

%--------------------------------------------------------
% GARCH(1,1)
%--------------------------------------------------------
Parameters_GARCH_1_1 = GARCH_1_1_calibration(updated_rets, GARCH_1_1_Params);

LogLF_G(x)=...
-GARCH_1_1_Maxlikelihood(updated_rets, Parameters_GARCH_1_1);

%--------------------------------------------------------
% Likelihood Ratio
%--------------------------------------------------------
Log_L_Rat(x)=2*(LogLF_G_Lag_Vol(x)-LogLF_G(x));

end

%--------------------------------------------------------
% Calculates the average values of the parameters
average_parameter_one=mean(Parameter_One);
average_parameter_two=mean(Parameter_Two);
average_parameter_three=mean(Parameter_Three);
average_parameter_four=mean(Parameter_Four);

average_LoglikelihoodFunc=mean(LogLF_G_Lag_Vol);

%--------------------------------------------------------
% Calculates the variances of the parameters
var_parameter_one=var(Parameter_One);
var_parameter_two=var(Parameter_Two);
var_parameter_three=var(Parameter_Three);
var_parameter_four=var(Parameter_Four);

var_LogLF=var(LogLF_G_Lag_Vol);

%--------------------------------------------------------
% Histograms
%--------------------------------------------------------

% Histogram of Alpha
figure(1);
hist(Parameter_Two,50);
```

```
title('Histogram of Alpha');

% Histogram of Beta
figure(2);
hist(Parameter_Three,50);
title('Histogram of Beta')

% Histogram of Gamma
figure(3);
hist(Parameter_Four,50);
title('Histogram of Gamma')

%--------------------------------------------------------
% Plot for the Returns and Inferred Volatility
%--------------------------------------------------------

%NOTE: Appropriate values need to be arranged first:

% 1. Forcasted Returns
smp_ret (1:Time,1) = updated_rets(1,1:Time);

% 2. Scaled Forecasted Returns
smp_ret_mean = smp_ret - mean(smp_ret);

% 3. Forecasted Volatility
w_var (1:Time,1) = w(1,1:Time);

%-------------------------------------------------------
% Plot the figure
figure(7);
plot(smp_ret_mean); hold on
plot(w_var); hold off
title('Aug-GARCH(1,1) with Lagged Volume - Scaled Returns and Inferred Volatility')
legend('Returns', 'Aug-Garch(1,1) lag vol - Volatility', 'Location', 'northwest')

%-------------------------------------------------------
% Plot the figure
figure(8);
plot(w_var); hold off
title('Aug-GARCH(1,1) with Lagged Volume - Volatility')
legend('Aug-Garch(1,1) lag vol - Volatility', 'Location', 'northwest')

%-------------------------------------------------------
% Histogram of Gamma
figure(9);
hist(Log_L_Rat,50);
title('Histogram of Loglikelihood Ratio')
```

Monte Carlo Simulations - Model 4: Aug-GARCH(1,1) model – News Volume $(news_t)$

```
% Aug_GARCH_MonteCarlo_news.m

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Code: Aug_GARCH_MonteCarlo_news.m            %
% Student Name: Jonathan Nolan                 %
% Student Number: 16071514                     %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%-----------------------------------------------
% Clear all work spaces and previous code

clc
clear

%-----------------------------------------------
% Load your data sets - Choose from 5 companies:

% 1. ASTRAZENECA
COMP = xlsread('astraftseN.xlsx',1, 'A:G','basic');
% 2. BP
%COMP = xlsread('bpftseN.xlsx',1, 'A:G','basic');
% 3. GLAXOSMITHKLINE
%COMP = xlsread('glaxoftseN.xlsx',1, 'A:G','basic');
% 4. TESCO
%COMP = xlsread('tescoftseN.xlsx',1, 'A:G','basic');
% 5. VODAFONE
%COMP = xlsread('vodaftseN.xlsx',1, 'A:G','basic');

%-----------------------------------------------
% Calculate the log returns:
ret = diff(log(COMP(:,3)));
ret(isnan(ret))=0;
ret(~isfinite(ret))=0;

%---------------------------------------------------------
fts_rets = diff(log(COMP(:,6)));
fts_rets(isnan(fts_rets))=0;
fts_rets(~isfinite(fts_rets))=0;
```

```
%-----------------------------------------------
Initial_1=9;
Initial_2=5;
Initial_3=5;

%Length of dataset
J = length(COMP);

%NEWS
news_vol = COMP(2:J,2);

%--------------------------------------------------------
% Aug-GARCH(1,1) Lagged Vol Params

Params=Aug_GARCH_1_1_runALL_news(Initial_1,Initial_2,Initial_3,ret,fts_rets,news_vo
l);

%--------------------------------------------------------
% GARCH(1,1) Parameters
%--------------------------------------------------------

GARCH_1_1_Params = GARCH_1_1_runALL(Initial_1,Initial_2, ret);

%--------------------------------------------------------
% Other Variables
%--------------------------------------------------------

Time = length(ret);
Num=1000;

for x=1:Num

%--------------------------------------------------------
w=Monte_Carlo_G_News_sim ( ret,fts_rets, news_vol, Time, Parameters);

for u=1:Time
%--------------------------------------------------------
ec(u) = w(u)*random('normal',0,1);

end
%--------------------------------------------------------
updated_rets=ec;

%--------------------------------------------------------
% Aug-GARCH(1,1)- News
%--------------------------------------------------------

Param_AugG_News =...
```

```
Aug_GARCH_1_1_calibration_news (updated_rets, fts_rets, news_vol, Parameters);

Parameter_One(x)=Param_AugG_News(1);
Parameter_Two(x)=Param_AugG_News(2);
Parameter_Three(x)=Param_AugG_News(3);
Parameter_Four(x)=Param_AugG_News(4);
Parameter_Five(x) = Param_AugG_News(5);
Parameter_Six(x) = Param_AugG_News(6);

LogLF_G_News(x)=...
-Aug_GARCH_Maxlikelihood_news(updated_rets,fts_rets, news_vol, Param_AugG_News);

%--------------------------------------------------------
% GARCH(1,1)
%--------------------------------------------------------
Parameters_GARCH_1_1 = GARCH_1_1_calibration(updated_rets, GARCH_1_1_Params);

LogLF_G(x)=...
-GARCH_1_1_Maxlikelihood(updated_rets, Parameters_GARCH_1_1);

%--------------------------------------------------------
% Likelihood Ratio
%--------------------------------------------------------
LL_Ratio(x)=2*(LogLF_G_News(x)-LogLF_G(x));

end

%--------------------------------------------------------
% Calculates the mean values of the parameters
average_parameter_one=mean(Parameter_One);
average_parameter_two=mean(Parameter_Two);
average_parameter_three=mean(Parameter_Three);
average_parameter_four=mean(Parameter_Four);
average_parameter_five = mean(Parameter_Five);
average_parameter_six = mean(Parameter_Six);

average_LoglikelihoodFunc=mean(LogLF_G_News);

%--------------------------------------------------------
% Calculates the variances of the parameters
var_parameter_one=var(Parameter_One);
var_parameter_two=var(Parameter_Two);
var_parameter_three=var(Parameter_Three);
var_parameter_four=var(Parameter_Four);
var_parameter_five = var(Parameter_Five);
var_parameter_six = var(Parameter_Six);

var_LogLF=var(LogLF_G_News);
```

```
%---------------------------------------------------------
% Histograms
%---------------------------------------------------------

% Histogram of Alpha
figure(1);
hist(Parameter_Two,50);
title('Histogram of Alpha');

% Histogram of Beta
figure(2);
hist(Parameter_Three,50);
title('Histogram of Beta')

% Histogram of Gamma
figure(3);
hist(Parameter_Four,50);
title('Histogram of Gamma')

%---------------------------------------------------------
% Plot for the Returns and Inferred Volatility
%---------------------------------------------------------

% 1. Forcasted Returns
smp_ret (1:Time,1) = updated_rets(1,1:Time);

% 2. Scaled Forecasted Returns
smp_ret_mean = smp_ret - mean(smp_ret);

% 3. Forecasted Volatility
w_var (1:Time,1) = w(1,1:Time);

%--------------------------------------------------------
% Plot the figure
figure(7);
plot(smp_ret_mean); hold on
plot(w_var); hold off
title('Aug-GARCH(1,1) with News - Scaled Returns and Inferred Volatility')
legend('Returns', 'Garch Vol', 'Location', 'northwest')

%--------------------------------------------------------
% Plot the figure
figure(8);
plot(w_var); hold off
title('Aug-GARCH(1,1) with News - Volatility')
legend('Garch Vol', 'Location', 'northwest')
```

```
%-------------------------------------------------------
% Plot the Histogram of LR
figure(9);
hist(LL_Ratio,50);
title('Histogram of LR');
```

Monte Carlo Simulations - Model 5: Aug-GARCH(1,1) model – Lagged News ($\boldsymbol{news_{t-1}}$)

```
% Aug_GARCH_MonteCarlo_news.m

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Code: Aug_GARCH_MonteCarlo_news.m            %
% Student Name: Jonathan Nolan                 %
% Student Number: 16071514                     %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%-----------------------------------------------
% Clear all work spaces and previous code

clc
clear

%-----------------------------------------------
% Load your data sets - Choose from 5 companies:

% 1. ASTRAZENECA
COMP = xlsread('astraftseN.xlsx',1, 'A:G','basic');
% 2. BP
%COMP = xlsread('bpftseN.xlsx',1, 'A:G','basic');
% 3. GLAXOSMITHKLINE
%COMP = xlsread('glaxoftseN.xlsx',1, 'A:G','basic');
% 4. TESCO
%COMP = xlsread('tescoftseN.xlsx',1, 'A:G','basic');
% 5. VODAFONE
%COMP = xlsread('vodaftseN.xlsx',1, 'A:G','basic');

%-----------------------------------------------
% Calculate the log returns:
ret = diff(log(COMP(:,3)));
ret(isnan(ret))=0;
ret(~isfinite(ret))=0;
```

```
%-----------------------------------------------
fts_rets = diff(log(COMP(:,6)));
fts_rets(isnan(fts_rets))=0;
fts_rets(~isfinite(fts_rets))=0;

%-----------------------------------------------
% Initial Parameters
% NOTE: These will be used to find initial points of the search


Initial_1=9;
Initial_2=5;
Initial_3=5;

%Length of dataset
J = length(COMP);

%NEWS
news_vol = COMP(2:J,2);

%--------------------------------------------------------
% Aug-GARCH(1,1) Lagged News Params

Params=Aug_GARCH_1_1_runALL_n_lag(Initial_1,Initial_2,Initial_3,ret,fts_rets,news_v
ol);

%--------------------------------------------------------
% GARCH(1,1) Parameters
%--------------------------------------------------------

GARCH_1_1_Params = GARCH_1_1_runALL(Initial_1,Initial_2, ret);

%--------------------------------------------------------
% Other Variables
%--------------------------------------------------------

Time = length(ret);
Num=1000;

for x=1:Num

%--------------------------------------------------------
w=Monte_Carlo_G_News_sim ( ret,fts_rets, new_vol, Time, Params);

for u=1:Time
%--------------------------------------------------------
ec(u) = w(u)*random('normal',0,1);
```

```
end
%-------------------------------------------------------
updated_rets=ec;

%-------------------------------------------------------
% Aug-GARCH(1,1)- Lag News
%-------------------------------------------------------

Param_AugG_News_Lag =...
Aug_GARCH_1_1_calibration_news_lag (updated_rets, fts_rets, news_vol, Parameters);

Parameter_One(x)=Param_AugG_News_Lag(1);
Parameter_Two(x)=Param_AugG_News_Lag(2);
Parameter_Three(x)=Param_AugG_News_Lag(3);
Parameter_Four(x)=Param_AugG_News_Lag(4);
Parameter_Five(x) = Param_AugG_News_Lag(5);
Parameter_Six(x) = Param_AugG_News_Lag(6);

LogLF_G_News_Lag(x)=...
-Aug_GARCH_Maxlikelihood_news_lag(updated_rets,fts_rets, news_vol,
Param_AugG_News_Lag);

%-------------------------------------------------------
% GARCH(1,1)
%-------------------------------------------------------
Parameters_GARCH_1_1 = GARCH_1_1_calibration(updated_rets, GARCH_1_1_Params);
LogLF_G(x)=...
-GARCH_1_1_Maxlikelihood(updated_rets, Parameters_GARCH_1_1);

%-------------------------------------------------------
% Likelihood Ratio
%-------------------------------------------------------
LL_Ratio(x)=2*(LogLF_G_News_Lag(x)-LogLF_G(x));
end

%-------------------------------------------------------
% Calculates the average values of the parameters
average_parameter_one = mean(Parameter_One);
average_parameter_two = mean(Parameter_Two);
average_parameter_three = mean(Parameter_Three);
average_parameter_four= mean(Parameter_Four);
average_parameter_five = mean(Parameter_Five);
average_parameter_six = mean(Parameter_Six);

average_LoglikelihoodFunc=mean(LogLF_G_News_Lag);

%-------------------------------------------------------
```

```
% Calculates the variances of the parameters
var_parameter_one = var(Parameter_One);
var_parameter_two = var(Parameter_Two);
var_parameter_three = var(Parameter_Three);
var_parameter_four = var(Parameter_Four);
var_parameter_five = var(Parameter_Five);
var_parameter_six = var(Parameter_Six);

var_LogLF=var(LogLF_G_News_Lag);

%--------------------------------------------------------
% Histograms
%--------------------------------------------------------
% Histogram of Alpha
figure(1);
hist(Parameter_Two,50);
title('Histogram of Alpha');

% Histogram of Beta
figure(2);
hist(Parameter_Three,50);
title('Histogram of Beta')

% Histogram of Gamma
figure(3);
hist(Parameter_Four,50);
title('Histogram of Gamma')

%--------------------------------------------------------
% Plot for the Returns and Inferred Volatility
%--------------------------------------------------------
%NOTE: Appropriate values need to be arranged first:
% 1. Forcasted Returns
smp_ret (1:Time,1) = updated_rets(1,1:Time);

% 2. Scaled Forecasted Returns
smp_ret_mean = smp_ret - mean(smp_ret);

% 3. Forecasted Volatility
w_var (1:Time,1) = w(1,1:Time);

%-------------------------------------------------------
% Plot the figure
figure(7);
plot(smp_ret_mean); hold on
plot(w_var); hold off
title('Aug-GARCH(1,1) with Lagged News - Scaled Returns and Inferred Volatility')
legend('Returns', 'Garch Vol', 'Location', 'northwest')
```

```
%-------------------------------------------------------
% Plot the figure
figure(8);
plot(w_var); hold off
title('Aug-GARCH(1,1) with Lagged News - Volatility')
legend('Garch Vol', 'Location', 'northwest')

%-------------------------------------------------------
% Plot the Histogram of LR
figure(9);
hist(LL_Ratio,50);
title('Histogram of LR');
```

Bibliography

Ahern, K.R., Sosyura, D. (2014) ‘Who Writes the News? Corporate Press Releases during Merger Negotiations’, *Journal of Finance*, 69(1), 241–291.

Alfano, S.J., Feuerriegel, S., Neumann, D. (2015) ‘Is News Sentiment More Than Just Noise?’, *Ecis*, (2015), 0–16.

Antweiler, W., Frank, M.Z. (2004) ‘Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards’, *The Journal of Finance*, 59(3), 1259–1294.

Bachelier, L., Davis, M., Etheridge, A. (2006) *Louis Bachelier’s Theory of Speculation : The Origins of Modern Finance* [online], Princeton University Press.

Banz, R.W. (1981) ‘The relationship between return and market value of common stocks’, *Journal of Financial Economics*, 9(1), 3–18.

Bettman, J.L., Hallett, A.G., Sault, S. (2010) ‘Rumortrage: Can Investors Profit on Takeover Rumors on Internet Stock Message Boards?’, *SSRN Electronic Journal*.

Bollen, J., Mao, H., Zeng, X.-J. (2011) Twitter Mood Predicts The Stock Market | MIT Technology Review [online], *Journal of Computational Science*.

Bollerslev, T. (1986) *GENERALIZED AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTICITY*, Journal of Econometrics.

De Bondt, W., Thaler, R. (1985) ‘Does the stock market overreact?’, *The Journal of Finance*, 40(3), 793–805.

Braun, P.A., Nelson, D.B., Sunier, A.M. (1995) ‘Good News, Bad News, Volatility, and Betas’, *Source: The Journal of Finance THE JOURNAL OF FINANCE*, 50(5), 1575–1603.

Brooks, C. (1998) ‘Predicting stock index volatility: can market volume help?’, *Journal of Forecasting*, 17(1), 59–80.

Brown, K.C., Harlow, W.V., Tinic, S.M. (1988) ‘Risk aversion, uncertain information, and market efficiency’, *Journal of Financial Economics*, 22(2), 355–385.

Bushee, B.J., Core, J.E., Guay, W.R., Hamm, S.J.W. (2010) 'The Role of the Business Press as an Information Intermediary', *Journal of Accounting Research*, 48(1), 1–19.

Chan, N.H., Ling, S. (2006) 'Empirical Likelihood for Garch Models', 22(3), 403–428.

Chen, H., De, P., Hu, Y., Hwang, B.-H., Faccio, M., Kim, S., Liberti, J., Lou, D., Loughran, T., Polak, I., Xu, J. (2014) 'Wisdom of Crowds: The Value of Stock Opinions Trandmitted Through SocialL MEDIA', *Review of Financial Studies*, 27(5), 1367–1403.

Core, J.E., Guay, W.R., Larcker, D.F. (2008) 'The Power of the Pen and Executive Compensation', *Journal of Financial Economics*, 88(1), 1–25.

Das, S.R. (2012) 'News Analytics: Framework, Techniques, and Metrics', in *The Handbook of News Analytics in Finance*, 41–71.

Das, S.R., Chen, M.Y. (2007) 'Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web', *Management Science*, 53(9).

Edmans, A., Goncalves-Pinto, L., Wang, Y., Xu, M. (2014) 'Strategic news releases in equity vesting months', 53(9), 1689–1699.

Fama, E.F. (1970) 'Efficient Capital Markets: A review of Theory and Empirical Work', *The Journal of Finance1*, 25(2), 383–417.

Groß-Klußmann, A., Hautsch, N. (2011) 'When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions'.

Hansen, P.R., Lunde, A. (2005) 'A Forecast Comparison of Volatility Models : Does Anything Beat a GARCH (1 , 1)?', *Journal of Applied Econometics*, 20(7), 873–889.

Harris, L. (1987) 'Transaction Data Tests of the Mixture of Distributions Hypothesis', *Journal of Financial and Quantitative Analysis*, 22(2), 127–141.

Henry, E. (2006) 'Market Reaction to Verbal Components of Earnings Press Releases: Event Study Using a Predictive Algorithm', *Journal of Emerging Technologies in …*, 3, 1–19.

Henry, E. (2008) 'Are Investors Influenced By How Earnings Press Releases Are Written?', *Journal of Business Communication*, 45(4), 363–407.

Janssen, G. (2004) 'Public information arrival and volatility persistence in financial markets', *The European Journal of Finance*, 10(3), 177–197.

Jorion, P. (1988) ‘On Jump Processes in the Foreign Exchange and Stock Markets’, *Review of Financial Studies*, 1(4), 427–445.

Kantardzic, M. (2011) ‘Data mining’, in *SpringerReference*, 534.

Kim, S.-H., Kim, D. (2014) ‘Investor sentiment from internet message postings and the predictability of stock returns’, *Journal of Economic Behavior & Organization*, 107, 708–729.

Kogan, S., Moskowitz, T.J., Niessner, M. (2017) *Fake News in Financial Markets*.

Lamoureux, C.G., Lastrapes, W.D. (1990) *American Finance Association Heteroskedasticity in Stock Return Data: Volume versus GARCH Effects*, Source: The Journal of Finance.

Li, Q., Wang, T., Li, P., Liu, L., Gong, Q., Chen, Y. (2014) ‘The effect of news and public mood on stock movements’, *Information Sciences*, 278, 826–840.

Li, X., Wang, C., Dong, J., Wang, F., Deng, X., Zhu, S. (2011) ‘Improving Stock Market Prediction by Integrating Both Market News and Stock Prices’, Springer, Berlin, Heidelberg, 279–293.

Loughran, T., Mcdonald, B. (2011) ‘When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks’, *THE JOURNAL OF FINANCE @BULLET*, LXVI(1).

Maat, H.P. (2007) ‘How Promotional Language in Press Releases Is Dealt With by Journalists: Genre Mixing or Genre Conflict?’, *Journal of Business Communication*, 44(1), 59–95.

Malkiel, B.G. (2003) ‘The Efficient Market Hypothesis and Its Critics’, *Journal of Economic Perspectives—Volume*, 17(1—Winter), 59–82.

Mitra, G., Mitra, L. (2011) *The Handbook of News Analytics in Finance*, Wiley.

Nann, S., Krauss, J., Schoder, D. (2013) ‘Predictive Analytics On Public Data - The Case Of Stock Markets’, in *Proceedings of the 21st European Conference on Information Systems*, 1–12.

Pantzalis, C., Stangeland, D.A., Turtle, H.J. (2000) ‘Political elections and the resolution of uncertainty: The international evidence’, *Journal of Banking & Finance*, 24(10), 1575–1604.

Park, J., Konana, P., Gu, B., Kumar, A., Raghunathan, R. (2013) ‘Information Valuation and Confirmation Bias in Virtual Communities: Evidence from Stock Message Boards’, *Information Systems Research*, 24(4), 1050–1067.

Peck, M., Hall, A. (2013) ‘Effective use of Press Releases in the Marketing of Academic Journals’, *Editors’ Bulletin*, 9(3), 60–66.

Pennebaker, J.W., Boyd, R.L., Jordan, K., Blackburn, K. (2015) ‘The Development and Psychometric Properties of LIWC2015’.

Sabherwal, S., Sarkar, S.K., Zhang, Y. (2011) ‘Do Internet Stock Message Boards Influence Trading? Evidence from Heavily Discussed Stocks with No Fundamental News’, *Journal of Business Finance & Accounting*, 38(9–10), 1209–1237.

Shen, D., Li, X., Zhang, W. (2017) ‘Baidu news information flow and return volatility: Evidence for the Sequential Information Arrival Hypothesis’, *Economic Modelling*, 69(May 2017), 127–133.

Si, J., Mukherjee, A., Liu, B., Li, Q., Li, H., Deng, X. (2013) ‘Exploiting Topic based Twitter Sentiment for Stock Prediction’, *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, (2011), 24–29.

Sprenger, T.O., Tumasjan, A., Sandner, P.G., Welpe, I.M. (2014) ‘Tweets and Trades: the Information Content of Stock Microblogs’, *European Financial Management*, 20(5), 926–957.

Tauchen, G.E., Pitts’, M. (1983) *THE PRICE VARIABILITY-VOLUME RELATIONSHIP ON SPECULATIVE MARKETS*, Econometrica.

Tetlock, P.C. (2007a) ‘Giving Content to Investor Sentiment : The Role of Media in the Stock Market’, *The Journal of Finance*, 62(3), 1139–1168.

Tetlock, P.C. (2007b) ‘Giving Content to Investor Sentiment: The Role of Media in the Stock Market Giving Content to Investor Sentiment: The Role of Media in the Stock Market’, *The Journal of Finance*.

Wang, G., Wang, T., Wang, B., Sambasivan, D., Zhang, Z., Zheng, H., Zhao, B.Y. (2014) ‘Crowds on Wall Street: Extracting Value from Social Investing Platforms’.

Wisniewski, T.P., Lambe, B. (2013) ‘The role of media in the credit crunch: The case of the banking sector’, *Journal of Economic Behavior and Organization*, 85(1), 163–175.

Yu, S., Rentzler, J., Tandon, K. (2009) ‘Reexamining the uncertain information hypothesis on the S&P 500 Index and SPDRs’.

Zarowin, P. (1990) ‘Size, Seasonality, and Stock Market Overreaction’, *Journal of Financial and Quantitative Analysis Mar*, 25(1).

Zhang, J.L., Härdle, W.K., Chen, C.Y., Bommes, E., Ardle, W.K. (2015) ‘Distillation of News Flow Into Analysis of Stock Reactions’, *Journal of Business & Economic Statistics*, 34(4), 547–563.

Zhang, X., Fuehres, H., Gloor, P.A. (2012) ‘Predicting asset value through twitter buzz’, in *Advances in Intelligent and Soft Computing*, 23–34.

Zhang, Y., Swanson, P.E. (2009) ‘Are day traders bias free?-evidence from internet stock message boards’, *Journal of Economics and Finance*, 34(1), 96–112.